

Epperson 09/526,106

=> d que 12

L1 585 SEA FILE=REGISTRY ELDLNSG[KE] ILESFR[SP] EERFPMMS.{0-}PNDERDTT[TM]
]PVAMATT/SQSP

L2 0 SEA FILE=REGISTRY L1 AND SQL<=207

=> d que 13

L3 0 SEA FILE=REGISTRY GEILESFRPEERFP.{0-}DTTMPVAM|GKILESFRSEERFP.{0-}
-}DTTMPVAM|GKILESFRPEERFP.{0-}DTTMPVAM|GEILESFRSEERFP.{0-}DTTMP
VAM|GEILESFRPEERFP.{0-}DTTMPVAM|GKILESFRSEERFP.{0-}DTTMPVAM|GEI
LESFRSEERFP.{0-}DTTMPVAM/SQSP

=> d que 114

L1 585 SEA FILE=REGISTRY ELDLNSG[KE] ILESFR[SP] EERFPMMS.{0-}PNDERDTT[TM]
]PVAMATT/SQSP

L4 78 SEA FILE=HCAPLUS L1

L5 2 SEA FILE=HCAPLUS L4 AND COMPLEMENTATION

L6 156 SEA BALINT R?/AU

L7 137 SEA HER J?/AU

L8 276 SEA L6 OR L7

L9 9 SEA L8 AND COMPLEMENTATION

L10 55396 SEA BETA(A) LACTAMAS?

L11 30 SEA L10(5A) COMPLEMENTATION

L12 23 SEA L11 NOT L9

L13 7 SEA L12 AND FRAGMENT?

L14 9 DUP REM L5 L9 L13 (9 DUPLICATES REMOVED)

=> d ibib abs 114 1-9

L14 ANSWER 1 OF 9 HCAPLUS. COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2003:551735 HCAPLUS

DOCUMENT NUMBER: 139:112705

TITLE: Methods for detection of molecular and protein
interactions by reporter subunit
complementation and its use in functional
genomics and drug screening

INVENTOR(S): Blau, Helen M.; Balint, Robert F.; Wehrman,
Thomas S.; Her, Jeng-horng

PATENT ASSIGNEE(S): The Board of Trustees of the Leland Stanford Junior
University, USA

SOURCE: PCT Int. Appl., 63 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2003058197	A2	20030717	WO 2002-US41587	20021226
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
PT, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,
MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

US 2001-344757P P 20011226

AB Methods and compns. for detecting mol. interactions, particularly, protein-protein interactions, using at least two inactive, weakly-complementing ss-lactamase fragments are provided. The invention allows detection of such interactions in eukaryotic and mammalian cells, in situ or in vitro. Detection of mol. interactions in mammalian cells is not limited to the nuclear compartment, but can be accomplished in the cytoplasm, cell surface, organelles, or between these entities. Methods provided utilize novel compns. comprising fusion proteins between mols. of interest and inactive, weakly-complementing-ss-lactamase fragments. Assocn. of the mols. of interest brings the corresponding complementary ss-lactamase fragments into close enough proximity for **complementation** to occur and ss-lactamase activity to be obsd. The invention is useful in the study of protein-protein interactions, functional genomics, agonist and antagonist screening and drug discovery.

L14 ANSWER 2 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2003:696387 HCAPLUS

TITLE: Interaction-activated, circularly permuted proteins

INVENTOR(S): Balint, Robert F.; Her, Jeng-Horng

PATENT ASSIGNEE(S): USA

SOURCE: U.S. Pat. Appl. Publ., 49 pp., Cont.-in-part of U.S. Ser. No. 526,106.

CODEN: USXXCO

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 3

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2003165825	A1	20030904	US 2001-764163	20010116
PRIORITY APPLN. INFO.:			US 2000-175968P	P 20000113
			US 2000-526106	A2 20000315

AB Interaction-activated circularly permuted proteins are disclosed that depend for their functional reassembly into the parent protein on the interaction of heterologous polypeptides or other mols. which have been genetically or chem. conjugated to the break-point termini of engineered enzymes. In addn., methods are provided for identifying circularly permuted marker proteins that will optimally reassemble into a functional parent protein, and which are dependent on the assocn. of heterologous interactor domains. The invention is exemplified by circular permutations of a Class A .beta.-lactamase (TEM-1 of E. coli). Circularly permuted marker proteins that comprise mol. interaction-dependent enzymes particularly find use in (1) cell-based sensors for activation or inhibition of metabolic or signal transduction pathways for high-efficiency, (2) high-throughput screening for agonists/antagonists of the target pathway and in high-throughput mapping of pair-wise protein-protein interactions within and between the proteomes of cells, tissues, and pathogenic organisms, and in (3) cell-based screens for high-throughput selection of inhibitors of any protein-protein interaction.

L14 ANSWER 3 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:353582 HCAPLUS

THIS PAGE BLANK (USPTO)

DOCUMENT NUMBER: 136:352286
 TITLE: Affinity maturation by competitive selection
 INVENTOR(S): **Balint, Robert F.; Her, Jeng-Horng**
 ; Larrick, James W.
 PATENT ASSIGNEE(S): Horizon Biotechnologies, Inc., USA
 SOURCE: PCT Int. Appl., 45 pp.
 CODEN: FIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 1
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002036738	A2	20020510	WO 2001-US45371	20011030
WO 2002036738	A3	20020704		
W:	AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG			
AU 2002027112	A5	20020515	AU 2002-27112	20011030
US 2002155502	A1	20021024	US 2001-999413	20011030
EP 1330551	A2	20030730	EP 2001-992763	20011030

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR
 PRIORITY APPLN. INFO.: US 2000-245039P P 20001030
 WO 2001-US45371 W 20011030

AB The present invention provides a method of selecting binding pair members with enhanced binding affinity for the cognate binding partner relative to a ref. binding pair member. In particular, the invention provides methods of selecting antibodies with enhanced affinity for an antigen relative to a ref. antibody. This process, "affinity maturation", thereby provides antibodies with superior binding capabilities. Examples show competitive detn. of **complementation** groups among thioredoxin-scaffolded peptides that bind human CD40 and affinity competition between an anti-CD40 antibody and a higher-affinity mutant of the same antibody.

L14 ANSWER 4 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER: 2002:266601 HCAPLUS
 DOCUMENT NUMBER: 137:29990
 TITLE: Protein-protein interactions monitored in mammalian cells via **complementation** of .beta.-lactamase enzyme fragments
 AUTHOR(S): Wehrman, Tom; Kleaveland, Benjamin; **Her, Jeng-Horng; Balint, Robert F.**; Blau, Helen M.
 CORPORATE SOURCE: Baxter Laboratory for Genetic Pharmacology, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, 94305, USA
 SOURCE: Proceedings of the National Academy of Sciences of the United States of America (2002), 99(6), 3469-3474
 CODEN: PNASA6; ISSN: 0027-8424
 PUBLISHER: National Academy of Sciences
 DOCUMENT TYPE: Journal

THIS PAGE BLANK (USP)

THIS PAGE BLANK (USP)

LANGUAGE: English

AB We have defined inactive .alpha. and .omega. fragments of .beta.-lactamase that can complement to form a functional enzyme in both bacteria and mammalian cells, serving as a readout for the interaction of proteins fused to the fragments. Crit. to this advance was the identification of a tripeptide, Asn-Gly-Arg, which when juxtaposed at the carboxyl terminus of the .alpha. fragment increased complemented enzyme activity by up to 4 orders of magnitude. .beta.-Lactamase is well suited to monitoring constitutive and inducible protein interactions because it is small (29 kDa), monomeric, and assayable with a fluorescent cell-permeable substrate. The negligible background, the magnitude of induced signal caused by enzymic amplification, and detection of signal within minutes are unparalleled in mammalian protein interaction detection systems published to date.

REFERENCE COUNT: 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 5 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:415162 HCAPLUS

DOCUMENT NUMBER: 137:137336

TITLE: A glutamine transport gene, glnQ, is required for fibronectin adherence and virulence of group B streptococci

AUTHOR(S): Tamura, Glen S.; Nittayajarn, Aphakorn; Schoentag, Deborah L.

CORPORATE SOURCE: Children's Hospital and Regional Medical Center and the University of Washington, Seattle, WA, 98105, USA

SOURCE: Infection and Immunity (2002), 70(6), 2877-2885

CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Group B streptococci (GBS) are a leading cause of neonatal sepsis and meningitis. GBS adhere to fibronectin when it is attached to a solid phase. We isolated a Tn917 transposon mutant, COH1-GT1, which shows decreased adherence to fibronectin. COH1-GT1 also shows decreased adherence to and invasion of respiratory epithelial cells in vitro and decreased virulence in vivo. COH1-GT1 contains a Tn917 insertion in a homolog of glnQ, a gene from Escherichia coli which is required for glutamine transport and codes for a cytoplasmic ATP-binding cassette protein. To confirm that the decreased fibronectin adherence of COH1-GT1 was due to the mutation in glnQ, we constructed COH1-GT2, a strain with a nonpolar site-directed mutation in glnQ. COH1-GT2 showed decreased binding to fibronectin. We also demonstrated that **complementation** of glnQ in trans restored fibronectin adherence to COH1-GT1. COH1-GT1 shows decreased uptake of radiolabeled glutamine and is resistant to the toxic glutamine analog .gamma.-L-glutamylhydrazide, demonstrating that the glnQ gene is required for glutamine transport in GBS. GlnQ lacks a signal sequence and is a cytoplasmic protein in E. coli and thus is unlikely to act as a fibronectin adhesin. GlnQ is transcribed in an operon with a putative glutamine permease gene, glnP, which has a novel predicted structure contg. three distinct domains linked in a single gene. The first two domains are putative glutamine binding domains with homol. to the E. coli periplasmic glutamine binding gene glnH. The third is a putative permease domain with homol. to the E. coli glutamine permease gene glnP. RT-PCR anal. demonstrated that glnP and glnQ are contained within a single transcript. Transcription of scpB, encoding the only known fibronectin-binding adhesin of GBS, is unaffected. We speculate that glnQ may regulate expression of fibronectin adhesins by affecting

THIS PAGE BLANK (USP)

THIS PAGE BLANK (USPTO)

cytoplasmic glutamine levels and that regulation may be posttranscriptional.

REFERENCE COUNT: 33 THERE ARE 33 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 6 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 2

ACCESSION NUMBER: 2002:417208 HCAPLUS

DOCUMENT NUMBER: 137:105986

TITLE: **.beta.-Lactamase** protein
fragment complementation assays as
in vivo and in vitro sensors of protein-protein
interactions

AUTHOR(S): Galarneau, Andre; Primeau, Martin; Trudeau,
Louis-Eric; Michnick, Stephen W.

CORPORATE SOURCE: Departement de Biochimie, Universite de Montreal,
Montreal, QC, H3C 3J7, Can.

SOURCE: Nature Biotechnology (2002), 20(6), 619-622
CODEN: NABIF9; ISSN: 1087-0156

PUBLISHER: Nature America Inc.

DOCUMENT TYPE: Journal

LANGUAGE: English

AB We have previously described a strategy for detecting protein-protein interactions based on protein interaction-assisted folding of rationally designed **fragments** of enzymes. We call this strategy the protein **fragment** complementation assay (PCA). Here we describe PCAs based on the enzyme TEM-1 **.beta.-lactamase** (EC: 3.5.2.6), which include simple colorimetric in vitro assays using the cephalosporin nitrocefin and assays in intact cells using the fluorescent substrate CCF2/AM (ref. 6). Constitutive protein-protein interactions of the GCN4 leucine zippers and of apoptotic proteins Bcl2 and Bad, and the homodimerization of Smad3, were tested in an in vitro assay using cell lysates. With the same in vitro assay, we also demonstrate interactions of protein kinase PKB with substrate Bad. The in vitro assay is facile and amenable to high-throughput modes of screening with signal-to-background ratios in the range of 10:1 to 250:1, which is superior to other PCAs developed to date. Furthermore, we show that the in vitro assay can be used for quant. anal. of a small mol.-induced protein interaction, the rapamycin-induced interaction of FKBP and yeast FRB (the FKBP-rapamycin binding domain of TOR (target of rapamycin)). The assay reproduces the known dissocn. const. and no. of sites for this interaction. The combination of in vitro colorimetric and in vivo fluorescence assays of **.beta.-lactamase** in mammalian cells suggests a wide variety of sensitive and high-throughput large-scale applications, including in vitro protein array anal. of protein-protein or enzyme-protein interactions and in vivo applications such as clonal selection for cells expressing interacting protein partners.

REFERENCE COUNT: 20 THERE ARE 20 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 7 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2001:904573 HCAPLUS

DOCUMENT NUMBER: 136:50647

TITLE: A protein **fragment** complementation assay
(PCA) for the detection of protein-protein,
protein-small molecule, and protein nucleic acid
interactions based on the E. coli TEM-1
.beta.-lactamase

INVENTOR(S): Michnick, Stephen W.; Galarneau, Andre

PATENT ASSIGNEE(S): Odyssey Pharmaceuticals, Inc., USA

THIS PAGE BLANK (USPTO)

SOURCE: PCT Int. Appl., 29 pp.
 CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 1
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001094617	A2	20011213	WO 2001-US17886	20010601
WO 2001094617	A3	20030206		
W: AU, CA, JP				
RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR				
US 2003108869	A1	20030612	US 2001-870018	20010531
EP 1305627	A2	20030502	EP 2001-983266	20010601
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI, CY, TR				

PRIORITY APPLN. INFO.:
 US 2000-208485P P 20000602
 US 2001-870018 A 20010531
 WO 2001-US17886 W 20010601

AB The present invention relates generally to protein complementation assays (PCA) and more specifically to PCA assays based on the E. coli TEM-1 .beta.-lactamase for the detection of protein-protein, protein-small mol., and protein nucleic acid interactions. In the present invention, an in vitro colorimetric assay using the substrate, nitrocefin, and an in vivo fluorescence assay using the substrate, CCF2/AM, were disclosed in mammalian cells. The invention is also directed to pos. and neg. survival assays using cephalosporin-cytotoxic prodrug conjugates.

L14 ANSWER 8 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 3

ACCESSION NUMBER: 2000:842264 HCAPLUS
 DOCUMENT NUMBER: 134:13994
 TITLE: High-performance enzyme fragment
complementation systems for the identification of interacting proteins
 INVENTOR(S): Balint, Robert F.; Her, Jeng-Horng
 PATENT ASSIGNEE(S): Panorama Research, Inc., USA
 SOURCE: PCT Int. Appl., 94 pp.
 CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000071702	A1	20001130	WO 2000-US7108	20000316
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
EP 1183347	A1	20020306	EP 2000-946748	20000316
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,				

THIS PAGE BLANK (USPTO)

IE, SI, LT, LV, FI, RO
 JP 2003500051 T2 20030107 JP 2000-620079 20000316
 PRIORITY APPLN. INFO.: US 1999-135926P P 19990525
 US 2000-175968P P 20000113
 WO 2000-US7108 W 20000316

AB Fragment pairs of a class A .beta.-lactamase (TEM-1 of Escherichia coli) are disclosed that depend for their functional reassembly into the parent protein on the interaction of heterologous polypeptides or other mols. which have been genetically or chem. conjugated to the break-point termini of the fragment pairs. In addn., methods are provided for identifying fragment pairs that will optimally reassemble into a functional parent protein. Fragment pairs that comprise mol. interaction-dependent enzymes find use in (1) homogeneous assays and biosensors for any analyte having two or more independent binding sites, (2) tissue-localized activation of therapeutic and imaging reagents in vivo for early detection and treatment of cancer, chronic inflammation, atherosclerosis, amyloidosis, infection, transplant rejection, and other pathologies, (3) cell-based sensors for activation or inhibition of metabolic or signal transduction pathways for high-efficiency, high-throughput screening for agonists/antagonists of the target pathway, (4) high-throughput mapping of pair-wise protein-protein interactions within and between the proteomes of cells, tissues, and pathogenic organisms, (5) rapid selection of antibody fragments or other binding proteins which bind specifically to polypeptides of interest, (6) rapid antigen identification for anti-cell and anti-tissue antibodies, (7) rapid epitope identification for antibodies, (10) cell-based screens for high-throughput selection of inhibitors of any protein-protein interaction.

REFERENCE COUNT: 13 THERE ARE 13 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 9 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1992:525518 HCAPLUS

DOCUMENT NUMBER: 117:125518

TITLE: Nucleotide sequence of the .beta.-lactamase gene from Enterococcus faecalis HH22 and its similarity to staphylococcal .beta.-lactamase genes

AUTHOR(S): Zscheck, Karen K.; Murray, Barbara E.

CORPORATE SOURCE: Med. Sch., Univ. Texas, Houston, TX, 77030, USA

SOURCE: Antimicrobial Agents and Chemotherapy (1991), 35(9), 1736-40

CODEN: AMACQ; ISSN: 0066-4804

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The nucleotide sequence of the constitutively produced .beta.-lactamase (Bla) gene from E. faecalis HH22 was shown to be identical to the published sequences of 3 of 4 staphylococcal type A .beta.-lactamase genes; more differences were seen with the genes for staphylococcal type C and D enzymes. One hundred forty nucleotides upstream of the .beta.-lactamase start codon were detd. for an inducible staphylococcal .beta.-lactamase and were identical to those of th constitutively expressed enterococcal gene, indicating that the changes resulting in constitutive expression are not due to changes in the promoter or operator region. Moreover, complementation studies indicated that prodn. of the enterococcal enzyme could be repressed. The genes for the enterococcal Bla and an inducible staphylococcal Bla were each cloned into a shuttle vector and transformed into enterococcal and staphylococcal recipients. The major difference between the backgrounds of the 2 hosts was that more enzyme was produced by the staphylococcal host, regardless of the source of the gene. The location of the enzyme was found to be host dependent,

THIS PAGE BLANK (USPTO)

since cloned gene generated extracellular (free) enzyme in the staphylococcus and cell-bound enzyme in the enterococcus. On the basis of the identities of the enterococcal Bla and several staphylococcal Bla sequences, these data suggest the recent spread of .beta.-lactamase to enterococci and also suggest the loss of a functional repressor.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

```

XX AAB59052;
AC
XX
XX 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
XX
XX Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neutropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antiviral; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX MO20005173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-611515/58.
XX
XX N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX
XX Claim 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neutropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antiviral; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiac activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemias; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX
XX Sequence 94 AA;
XX
XX
XX Query Match 34.0%; Score 457; DB 21; Length 94;
XX Best Local Similarity 97.8%; Pred. No. 1.5e-39;
XX Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX
XX 75 NDVEXSPYTERKHLTDGTVRELCSAATYMSDNTAANLLTTTGGPKETAFIHNKGDHV 134
XX |
XX 1 NDVEXSPYTERKHLTDGTVRELCSAATYMSDNTAANLLTTTGGPKETAFIHNKGDHV 60
XX
XX 135 TRLDREPEPLNAIINDERDRTTPVAAAT 164
XX |
XX 61 TRLDREPEPLNAIINDERDRTTPVAAAT 90
XX

```

```

RESULT 9
AAP60628
ID AAP60628 standard; Protein; 159 AA.
XX
XX AAP60628;
AC
XX
XX 15-MAR-2003 (updated)
XX 17-JUN-1991 (first entry)
DE Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
XX
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX /label= beta-lactamase
XX Misc-difference 97.100
XX /label= adaptor
XX Protein 101..159
XX /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART ) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgaki H, Horioka A, Hiramatsu H, Koumoto S, Nishimura A;
XX Matsushiro S;
XX
XX WPI; 1986-015031/03.
XX
XX N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX and transformed cells contg. it.
XX
XX PS Disclosure: Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E.coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obtd. only
XX in small ants. from human urine.
XX See also AAN60628 and 30-32.
XX
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX
XX Sequence 159 AA;
XX
XX
XX Query Match 29.4%; Score 396; DB 7; Length 159;
XX Best Local Similarity 67.5%; Pred. No. 7.2e-33;
XX Matches 85; Conservative 9; Mismatches 20; Indels 12; Gaps 2;
XX
XX
XX 1 HPEITLVKVADEADQAGRVGYIELDINSGEILSFSEERFPFMSTFKVLCGAVLSRID 60
XX |
XX 24 HPEITLVKVADEADQAGRVGYIELDINSKILLESFRPERFPFMSTFKVLCGAVLSRID 83
XX
XX 61 AQEQQLGRIRIHSQNDLVYS-----PTEKHLTDGTVRELCSAATYMSDNTAANT 112
XX |
XX 84 AQEQQLGRIRIHSQNDLVESAKRNSDSECPISH---DGYCLHDCVCMYIEALDRYACNC 139
XX
XX 113 LLTITG 118
XX :: |
XX

```

Db 140 VGVIG 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein; 101 AA.
XX
AC ABG27935;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27926.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS92122.
XX

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 58294; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

XX Sequence 101 AA;
SQ

Query Match 27.2%; Score 366; DB 22; Length 101;
Best Local Similarity 98.6%; Pred. No. 4.9e-30;
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 154 DTTTPVAMATTIRKLTLGELLTLASROQLIDMWEADKAVAGPLLSALPAGWFIADKSGAG 213
DB 10 DTTTPVAMATTIRKLTLGELLTLASROQLIDMWEADKAVAGPLLSALPAGWFIADKSGAG 69

QY 214 ERGSRGIIAALGPD 227
DB 70 ERGSRGIIAALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein; 88 AA.
XX
AC ABG27919;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27910.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS92106.
XX

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 58278; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

XX Sequence 88 AA;
SQ

Query Match 24.2%; Score 325; DB 22; Length 88;
Best Local Similarity 80.0%; Pred. No. 7.3e-26;
Matches 68; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

QY 4 TLVVKRAEQLGARVYITLIDNSGRILSPFSEERFPKMSFKYLLGAVLSRIDAGQ 63
DB TLVVKRAEQLGARVYITLIDNSGRILSPFSEERFPKMSFKYLLGAVLSRIDAGQ 63

DB 1 TLVAVKDAEDQLGARVGYIELDLNSGRIIESFRPERPMPMSTPKVLLCGAVLSRIDAQ 60

OY 64 EQLGRRI-----HYSQNDLV 78
DB 61 EQLGRRIILRMWLTSHSQSKIL 85

RESULT 12

ID AAP60626 standard; Protein: 119 AA.

AC AAP60626;

DT 25-MAR-2003 (updated)
DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

KW Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
KW fusion protein; beta-lactamase.

OS Synthetic.

FT Key Location/Qualifiers

FT Protein 1..63 /label= beta-lactamase

FT Misc-difference 64..66 /label= adaptor

FT Protein 67..119 /label= beta-urogastrone

FT DE3523634-A.

PD 09-JAN-1986.

PF 02-JUL-1985; 85DE-3523634.

PR 02-JUL-1984; 84JP-0137691.

PA (EART) EARTH CHEM CO LTD.

PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Kounoto S, Nishimura A;
PI Matsushiro S;

DR WPI: 1986-015031/03.

DR N-PSDB; AAN60630.

PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PT and transformed cells contg. it.

PS Disclosure: Page 55; 92pp; German.

CC The fusion protein is less easily degraded by proteases and so

CC protects beta-urogastrone and beta-lactamase collects in the periplasm

CC of E.coli. It is therefore easy to collect and purify the product.

CC Beta-urogastrone is the hormone of the salivary glands which suppresses

CC stomach acid secretion and promotes cell growth, so is useful for

CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)

SQ Sequence 119 AA;

Query Match 14.2%; Score 191; DB 7; Length 119;
Best Local Similarity 95.0%; Pred. No. 9.6e-12;

Matches 38; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 HPETLVKVKDAEDQLGARVGYIELDLNSGRIIESFRSEK 40
DB 24 HPETLVKVKDAEDQLGARVGYIELDLNSGRIIESFRSEK 63

RESULT 13
ID AAY92783 standard; Protein: 51 AA.

AC AAY92783;

DT 29-AUG-2000 (first entry)

DE Vtgs-beta-lactamase fusion protein (partial).

KW Vtgs: vitellogenin; secretory signal sequence; gene expression;
KW oestrogen receptor binding protein; systemic circulation;
KW beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

FT Key Location/Qualifiers

FT Peptide 1..21 /label= Vtgs

FT Cleavage-site 15..16

FT Protein 22..51 /note= "beta-lactamase mature protein"

PN WO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

PA (UYSI-) UNIV SINGAPORE NAT.

PA (LAMT/) LAM T J.

PI Ding JL, Tan NS, Ho B;

DR WPI: 2000-365615/31.

DR N-PSDB; AAA28507.

PT Isolated nucleic acid for assaying for heterologous gene expression,
PT detecting presence of compound that binds to estrogen receptor or
PT producing desired protein from host cell comprises nucleotide sequence
PT encoding secretory signal sequence

PS Example 6; Fig 14A; 73pp; English.

CC A reporter beta-lactamase system that uses the Oreochromis aureus,
CC vitellogenin secretory sequence (Vtgs), designated PBADVgblactkana was

CC constructed. Vtgs and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal

CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound

CC that binds to an oestrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining

CC systemic circulation of a desired protein in a transgenic or chimeric
CC host organism.

SQ Sequence 51 AA;

Query Match 10.5%; Score 141; DB 21; Length 51;
Best Local Similarity 96.6%; Pred. No. 4.5e-07;

Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 PETLVKVKDAEDQLGARVGYIELDLNSGE 30
DB 23 PETLVKVKDAEDQLGARVGYIELDLNSGR 51

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345
Sequence: I HPEITLVKVKAEADQLGARVG.....TMDERNQDAEIGASLTKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/prodata/1/1aa/7CTUS.COMB.pep: *
6: /cgn2_6/prodata/1/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.5	5.5	124	US-08-311-731A-202	Sequence 202, App
2	72.5	5.4	157	US-09-252-991A-24241	Sequence 24241, A
3	71	5.3	159	US-08-991-890-4	Sequence 4, Appl
4	69	5.1	169	US-09-328-352-4491	Sequence 4491, Ap
5	67.5	5.0	170	US-08-858-207A-519	Sequence 519, App
6	66	4.9	181	US-08-482-142-195	Sequence 195, App
7	66	4.9	181	US-08-484-296-195	Sequence 195, App
8	66	4.9	181	US-08-484-296-195	Sequence 195, App
9	65.5	4.8	158	US-09-010-809-19	Sequence 19, Appl
10	65	4.8	150	US-09-239-909-2	Sequence 2, Appl
11	65	4.8	174	US-08-557-122A-6	Sequence 6, Appl
12	65	4.8	174	US-09-262-666-6	Sequence 6, Appl
13	65	4.8	200	US-08-557-122A-12	Sequence 12, Appl
14	65	4.8	200	US-09-262-666-12	Sequence 12, Appl
15	64.5	4.8	178	US-09-252-991A-29942	Sequence 29942, A
16	64	4.8	203	US-08-624-677A-2	Sequence 2, Appl
17	63.5	4.7	144	US-08-225-480-4	Sequence 4, Appl
18	63.5	4.7	144	US-09-118-445-4	Sequence 4, Appl
19	62	4.6	132	US-09-252-991A-22681	Sequence 22681, A
20	62	4.6	204	US-09-252-991A-17611	Sequence 17611, A
21	61.5	4.5	142	US-09-345-473E-31	Sequence 31, Appl
22	61	4.5	189	US-08-671-548C-48	Sequence 48, Appl
23	60.5	4.5	103	US-09-732-210-1282	Sequence 1282, Ap
24	60.5	4.5	167	US-08-690-849-2	Sequence 2, Appl
25	60.5	4.5	167	US-09-004-053-2	Sequence 2, Appl
26	60.5	4.5	198	US-09-413-814-87	Sequence 87, Appl
27	60	4.5	144	US-09-252-991A-25578	Sequence 25578, A

28	60	4.5	158	US-09-107-532A-4218	Sequence 4218, Ap
29	59.5	4.4	146	US-09-134-001C-5269	Sequence 5269, Ap
30	59.5	4.4	160	US-09-252-991A-24737	Sequence 24737, A
31	59	4.4	171	US-09-107-532A-3979	Sequence 3979, Ap
32	59	4.4	180	US-09-194-905-5	Sequence 5, Appl
33	59	4.4	199	US-09-252-991A-30363	Sequence 30363, A
34	58.5	4.3	203	US-09-252-991A-24921	Sequence 24921, A
35	58	4.3	72	US-08-858-207A-412	Sequence 412, App
36	58	4.3	136	US-09-252-991A-30622	Sequence 30622, A
37	58	4.3	168	US-08-451-947-6	Sequence 6, Appl
38	58	4.3	168	US-08-424-826A-6	Sequence 6, Appl
39	58	4.3	168	US-08-928-694-6	Sequence 6, Appl
40	58	4.3	168	US-08-450-842-6	Sequence 6, Appl
41	58	4.3	168	US-08-451-160-6	Sequence 6, Appl
42	58	4.3	168	PCT-US91-06950-6	Sequence 6, Appl
43	58	4.3	183	US-09-252-991A-31347	Sequence 31347, A
44	58	4.3	187	US-09-252-991A-29325	Sequence 29325, A
45	58	4.3	189	US-09-252-991A-17160	Sequence 17160, A

ALIGNMENTS

RESULT 1
US-08-311-731A-202
; Sequence 202, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MICROBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; NUMBER OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,731A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: C0044/7125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 202:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 124 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium leprae
; US-08-311-731A-202

Query Match 5.5%; Score 74.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;
QY 130 MGDHV--TRLDREWEPELN---EALPNDERT---TTPVAMATTLRKLTLGELLTLASRQ 180

```

Db      12 MGCGICMERGCRFTGTCPLRVVPGDESPITLDGRASPELITF--NLSPITMSPPS 69
QY      181 QLIDMNEA-DKVAGPFL-----RSALPAGWFIADSGAGSGRSRTIINALGPDGRSR 232
Db      70 RDDDWVEPFDALOGTAVFATGCKATWPAVGGI---GASTNGS-GILASLSPFROPAR 123

```

RESULT 2

```

US-09-352-991A-24241
; Sequence 24241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24241
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

```

```

Query Match      5.4%; Score 72.5; DB 4; Length 197;
Best Local Similarity 23.6%; Pred. No. 6.2;
Matches 37; Conservative 23; Mismatches 42; Indels 55; Gaps 7;

```

```

QY      94 VRELCSAITYMSDNTAANLLTTGGFKELTAFLHNGDHYTRLDRWEPELNEAIPNDER 153
Db      12 VAKSSAFTTSAASLASVPLPIATP---TSARFSAGASLT-----PEPVI 55
QY      154 DTTTPVA-MATTLKRLTGLTLTA-----SROQLDMKADKXVGP----- 194
Db      56 ATSPACRACRTRSLCSG--LARANSTPGSTSRRR--WSSNSISAPVSAGRSMPIP 110
QY      195 -----LRSALPAGWFIADSGAGSGERC 216
Db      111 SMADIAAVSTWSPVITFTAIAPAMHSATATASSRG 147

```

RESULT 3

```

US-08-991-890-4
; Sequence 4, Application US/08991890
; Patent No. 6114307
; GENERAL INFORMATION:
; APPLICANT: Jasper, Stephen R.
; APPLICANT: Sprugel, Katherine H.
; APPLICANT: Ren, Hong Ping
; APPLICANT: Humes, Jacqueline M.
; APPLICANT: Hoffman, Ross C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,890
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,003
; FILING DATE: December 16, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-08-991-890-4

```

```

Query Match      5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.5;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

```

```

QY      49 VLICGAVLSRIDGQ-----QLGRRIHYQN-----DLYEVSPTENH 87
Db      12 LILCGAVF--VSPQEIHAFFQGRKRRHHHGGSCAELRGCGPRGKRLSLCPMPERT 69
QY      88 LTGQMYRELCSAITYMSDNTAANLLTTGGFKELTAFLHNGDHYTRLDRWEPELNEA 147
Db      70 FT-----TTPGKWLES-GREKEMVSTNNKDGAL-----GTSSEF 105
QY      148 IPNDERDTTPVAMA--TTLKRL 169
Db      106 IPLSPELKRPPLSEGPSLKTI 128

```

RESULT 4

```

US-09-328-352-4491
; Sequence 4491, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4491
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

```

```

Query Match      5.1%; Score 69; DB 4; Length 169;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

```

```

QY      118 GPKRELTAFLHNGDHYTRLDRWEPELNEAIPNDEROT-----TPVAMATTLKRL 169
Db      5 GGENIMMLADON--RFDQWEQIVQYL--DRLTVRERIMVFTTFIVVVIVIGYS 58
QY      170 TGEILTLASROQ-----LIDMEADKVAGPILRSALPAGWFIADSGAGSG--GSRG 219
Db      59 LMKHSLAEQOQKRLNDKDLVMWQSNV-----IMKRNELDKSGRTGRTVVAQOQG 112
QY      220 IIALGPDGRKPSRIVYITTGSOATMDERNNOIAETASLIR 261

```

Db 113 LTVSSQNGEQLQIV--THQNVAILANFLTQLQMGSLIOK 152

RESULT 5
US-08-858-207A-519
; Sequence 519, Application US/08858207A
; Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Nicholas, Richard

APPLICANT: Stodola, Robert

TITLE OF INVENTION: No. 6348328el Compounds

NUMBER OF SEQUENCES: 552

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017670

FILING DATE: 14-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimm, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50475

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 519:

SEQUENCE CHARACTERISTICS:

LENGTH: 170 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6348328e

US-08-858-207A-519

Query Match 5.0%; Score 67.5; DB 4; Length 170;

Best Local Similarity 21.6%; Pred. No. 18;

Matches 46; Conservative 33; Mismatches 71; Indels 63; Gaps 8;

Db 15 LGARVGTIELDLSGELLESFRSEERPMWSTFKVLGCAVLSRIDAG--EOLGRRIHY 72
1 LKNIGVIVDLSIRELSPISCTKRSVMSRFK-----AAAFVADDSFIQELPQRYD- 54
73 SQNDLVEYSPYTEKHLTDGMTVRELCSAATMSDNTANLLTTIGSKRELTAFLNMGD 132
55 -----SPVSESGSFTSGORQLAFARTVASQ-----PKLLI----- 86
133 HYTRILRWPELNLAIINDEKDDTTTPVAMATTLKLLTGELTLASHOQLIDMWEADKVA 192
87 -----LDRATNIDSETSLV--QASLAKMGR-TTIAIHRLESTIODANCI- 131
193 GPLLRSAIPAGWFIADKSGAGSGRGIIAALG 225
132 -----YLDKGRILESTGTHELLAIG 152

RESULT 6
US-08-482-142-195
; Sequence 195, Application US/08482142
; Patent No. 5820862

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,142

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/445,307

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017,605

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 195:

SEQUENCE CHARACTERISTICS:

LENGTH: 181 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-482-142-195

Query Match 4.9%; Score 66; DB 2; Length 181;

Best Local Similarity 20.5%; Pred. No. 28;

Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5;

Db 4 TLVKVDAEDQAGARVY--TELDLSGELLESFRSEERPMWSTFKVLGCAVLSRIDA 61
17 THAKIRDVAAATESAVIATRTSLDSEQLVDERIGHNGVYQDSYTRYV----- 65
62 GQDOLGRRIHYSO-----NDLVEYSPYTEKHLTDGMTVRELCSAATMSDNTANLL 114
66 ARQSCRRPAAQGLAEVFEANQNTKAKIEIKASIDGLEV----- 105
115 TTIGSKRELTAFLNMGDHVTRLDRWPELNLAIIPND-----ENDTTTPVAM 161
106 --YIGIKIDDAFRHYDGRITIQDNGYOSINGNAPAEIDLRQRTVTPIRM 154

RESULT 7

US-08-478-572-195

; Sequence 195, Application US/08478572

; Patent No. 5968526

; GENERAL INFORMATION:

APPLICANT: Garman, Richard

```

: APPLICANT: Greenstein, Julia
: APPLICANT: Kuo, Wei-chang
: APPLICANT: Rogers, Bruce
: APPLICANT: Franzen, Henry
: APPLICANT: Chen, Xian
: APPLICANT: Evans, Sean
: APPLICANT: Shaked, Ze'ev
: TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESS: IMMULOGIC PHARMACEUTICAL CORPORATION
: STREET: 610 LINCOLN STREET
: CITY: WALTHAM
: STATE: MA
: COUNTRY: USA
: ZIP: 02154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII TEXT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/478,572
: FILING DATE: 07-June-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/445,307
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: CRAIG, ANNE I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: 017.605
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 466-6000
: TELEFAX: (617) 466-6040
: INFORMATION FOR SEQ ID NO: 195:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 181 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-478-572-195

Query Match      4.9%; Score 66; DB 2; Length 181;
Best Local Similarity 20.5%; Pred. No. 28;
Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5;

QY 4 TLKVKDAEDQLGARVGY--IEDLNSGELIESFRSEERPPMSTFKVLCGAVLSRIDA 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 17 THAKIRIVNATSEAYLAVRNTSLDSEQLVDEYIQHNGVQESYRYV----- 65
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 62 GQELGRIRHYSQ-----NDLVESPVTEKHLTDGMYRELCSAATMSDNTANLL 114
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 66 ARQSCRPRNAOQLAEVFEANONTKAKTEIKASIDGLEV----- 105
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 115 TTIGPRELTAFLHNGDHYTRLDRWEPELNEAIPND---ERDTTPVAM 161
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 106 --IIGKIDAFRRHYDGRITIQRDNGYOSINGNAPAEIDLROMKRTVTPIRM 154
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 8
: US-08-484-296-195
: Sequence 195, Application US/08484296
: Patent No. 6268491
: GENERAL INFORMATION:
: APPLICANT: Garman, Richard
: APPLICANT: Greenstein, Julia
: APPLICANT: Kuo, Wei-chang
: APPLICANT: Rogers, Bruce
: APPLICANT: Franzen, Henry
: APPLICANT: Chen, Xian
: APPLICANT: Evans, Sean
```

```

: APPLICANT: Shaked, Ze'ev
: TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESS: IMMULOGIC PHARMACEUTICAL CORPORATION
: STREET: 610 LINCOLN STREET
: CITY: WALTHAM
: STATE: MA
: COUNTRY: USA
: ZIP: 02154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII TEXT
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,296
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/445,307
: FILING DATE: 07 June 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CRAIG, ANNE I.
: REGISTRATION NUMBER: 32,976
: REFERENCE/DOCKET NUMBER: 017.605
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 466-6000
: TELEFAX: (617) 466-6040
: INFORMATION FOR SEQ ID NO: 195:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 181 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-484-296-195

Query Match      4.9%; Score 66; DB 3; Length 181;
Best Local Similarity 20.5%; Pred. No. 28;
Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5;

QY 4 TLKVKDAEDQLGARVGY--IEDLNSGELIESFRSEERPPMSTFKVLCGAVLSRIDA 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 17 THAKIRIVNATSEAYLAVRNTSLDSEQLVDEYIQHNGVQESYRYV----- 65
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 62 GQELGRIRHYSQ-----NDLVESPVTEKHLTDGMYRELCSAATMSDNTANLL 114
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 66 ARQSCRPRNAOQLAEVFEANONTKAKTEIKASIDGLEV----- 105
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 115 TTIGPRELTAFLHNGDHYTRLDRWEPELNEAIPND---ERDTTPVAM 161
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 106 --IIGKIDAFRRHYDGRITIQRDNGYOSINGNAPAEIDLROMKRTVTPIRM 154
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 9
: US-09-010-809-19
: Sequence 19, Application US/09010809B
: Patent No. 6090601
: GENERAL INFORMATION:
: APPLICANT: Gustafson, Claes
: APPLICANT: Betlach, Mary C.
: TITLE OF INVENTION: Epithione Polyketide Synthases and Encoding DNA
: FILE REFERENCE: 30062-20020.00
: CURRENT APPLICATION NUMBER: US/09/010,809B
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 158
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
```



```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-262-666-6

Query Match
Best Local Similarity 4.8%; Score 65; DB 4; Length 174;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KYDAEDQGLAR-VGYIELDLNSGELIESRSEERFPMMSTFKVLLCGAVLSRIDAGOEQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 67 KYEOATELKEKNIPLVKVDCTEEFALCRDQGEVGYPTLKIFRGL-----DAVKPY 117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 66 LGRR-----IHYS-QNDLVEYSPVTEKHLDGTFVRELCSAATMSDNTANILLTT 116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 118 QGAROTEALVSYWKSPLAVSPVTPENLEIKTKIVIGYIASDDOTANDIFFT 174

RESULT 13
US-08-557-122A-12
; Sequence 12, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Molland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980, 204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-122A-12

Query Match
Best Local Similarity 4.8%; Score 65; DB 2; Length 200;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KYDAEDQGLAR-VGYIELDLNSGELIESRSEERFPMMSTFKVLLCGAVLSRIDAGOEQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 43 KYEOATELKEKNIPLVKVDCTEEFALCRDQGEVGYPTLKIFRGL-----DAVKPY 93
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 66 LGRR-----IHYS-QNDLVEYSPVTEKHLDGTFVRELCSAATMSDNTANILLTT 116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 94 QGAROTEALVSYWKSPLAVSPVTPENLEIKTKIVIGYIASDDOTANDIFFT 150

RESULT 14
US-09-262-666-12
; Sequence 12, Application US/09262666
; Patent No. 6346244
```

```

; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Molland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980, 204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-262-666-12

Query Match
Best Local Similarity 4.8%; Score 65; DB 4; Length 200;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KYDAEDQGLAR-VGYIELDLNSGELIESRSEERFPMMSTFKVLLCGAVLSRIDAGOEQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 43 KYEOATELKEKNIPLVKVDCTEEFALCRDQGEVGYPTLKIFRGL-----DAVKPY 93
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 66 LGRR-----IHYS-QNDLVEYSPVTEKHLDGTFVRELCSAATMSDNTANILLTT 116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 94 QGAROTEALVSYWKSPLAVSPVTPENLEIKTKIVIGYIASDDOTANDIFFT 150

RESULT 15
US-09-252-991A-29942
; Sequence 29942, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubinfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29942
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
```



```

; LOCATION: (116)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match          4.8%; Score 64.5; DB 4; Length 178;
Best Local Similarity 26.8%; Pred. No. 41;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

OY      110 ANLLTTIG--GPKELTAF---LNNMGDHYTRLDWEPPELNEAIPNDRDPTTTPVAMA 162
        | : | | : | | | | : | | | | : | | | | : | | | | : | | | | : |
DB      25 ARVLVLELAQORRPGELAAHAGIAAHHHPGDHYAR-----AGEDLP---RDYDGLGAAA 74

OY      163 TTRKLLTGELT-----TLASRQQLIDWMEADKYAGPLLRSLALPAGWTADKSG 211
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      75 LRHHLYLAAGALOGHEHHEGIGDGTAGAEQAVVG--QDEEVAGSPXV--GIQAGLFLMVQGD 130

OY      212 AGEGRSGITIALGPDGK 229
        | : | | | |
DB      131 A-----LVGVVGGQAGK 141
```

Search completed: September 10, 2003, 12:31:05
Job time : 11.8571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(Without alignments)

2183.941 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345

Sequence: 1 HPELVKVADEQAGARVG.....TMDERKQIAETGASLIKRW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTOS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	457	34.0	94	US-10-102-806-760	Sequence 760, App
2	75	5.6	202	US-10-127-816-9	Sequence 9, Appl
3	75	5.6	202	US-10-142-717-12	Sequence 12, Appl
4	70.5	5.2	193	US-09-805-354-8	Sequence 8, Appl
5	70.5	5.2	193	US-10-144-259-8	Sequence 8, Appl
6	70	5.2	206	US-10-156-761-13867	Sequence 13867, A
7	69.5	5.2	184	US-10-101-464A-765	Sequence 765, App
8	69	5.1	149	US-10-156-761-8136	Sequence 8136, App
9	67.5	5.0	187	US-10-233-926-4	Sequence 4, Appl
10	67.5	5.0	187	US-09-948-018-8	Sequence 5880, Ap
11	67	5.0	178	US-10-156-761-8623	Sequence 8623, Ap
12	67	5.0	194	US-10-127-816-11	Sequence 11, Appl
13	66	4.9	202	US-09-815-242-5862	Sequence 5862, Ap
14	65.5	4.9	189	US-09-815-242-12979	Sequence 12979, A
15	65.5	4.9	189	US-09-815-242-12979	Sequence 12979, A

16	65.5	4.9	191	US-09-764-870-390	Sequence 390, App
17	65.5	4.9	191	US-10-125-540-390	Sequence 390, App
18	64.5	4.8	191	US-10-156-761-8978	Sequence 8978, Ap
19	64	4.8	162	US-09-738-626-4796	Sequence 4796, Ap
20	64	4.8	179	US-09-764-868-757	Sequence 757, App
21	64	4.8	179	US-10-106-658-4858	Sequence 4858, Ap
22	64	4.8	202	US-10-189-346-12	Sequence 12, Appl
23	63.5	4.7	88	US-10-100-282-6	Sequence 6, Appl
24	63.5	4.7	144	US-10-131-406-4	Sequence 4, Appl
25	63.5	4.7	202	US-10-156-761-13460	Sequence 13460, A
26	63.5	4.7	206	US-09-738-626-5425	Sequence 5425, Ap
27	63	4.7	197	US-09-803-661-5	Sequence 5, Appl
28	63	4.7	197	US-10-300-827-5	Sequence 5, Appl
29	62.5	4.6	160	US-09-882-227-230	Sequence 230, App
30	62.5	4.6	190	US-10-156-761-9507	Sequence 9507, Ap
31	62	4.6	174	US-09-864-761-35777	Sequence 35777, A
32	62	4.6	185	US-10-101-464A-616	Sequence 616, App
33	62	4.6	189	US-09-764-870-523	Sequence 523, App
34	62	4.6	189	US-10-125-540-523	Sequence 523, App
35	62	4.6	191	US-09-927-827-64	Sequence 64, Appl
36	61.5	4.6	142	US-09-862-027-31	Sequence 31, Appl
37	61.5	4.6	196	US-09-738-626-6624	Sequence 6624, Ap
38	60.5	4.5	202	US-10-189-346-16	Sequence 16, Appl
39	60	4.5	190	US-09-738-626-4637	Sequence 4637, Ap
40	59.5	4.4	182	US-10-156-761-12056	Sequence 12056, A
41	59.5	4.4	196	US-10-219-220-265	Sequence 265, App
42	59.5	4.4	206	US-09-741-669-461	Sequence 461, App
43	59.5	4.4	206	US-09-815-242-10335	Sequence 10335, A
44	59	4.4	111	US-09-867-550-1868	Sequence 1868, Ap
45	59	4.4	11	US-09-768-235B-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-10-102-806-760
Sequence 760, Application US/10102806
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA10391C1
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 760
LENGTH: 94
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (80)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 34.0% Score 457; DB 15; Length 94;
Best Local Similarity 97.8% Pred. No. 1.5e-35;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 75 NDVLESPVTEKHLTDQMTVRELCSAITSNDTANLLTTTGGPRELAFPHNGDHY 134
DB 1 NDVLESPVTEKHLTDQMTVRELCSAITSNDTANLLTTTGGPRELAFPHNGDHY 60


```

Query Match      5 28; Score 70; DB 15; Length 206;
Best Local Similarity 25.28; Pred. No. 44;
Matches 53; Conservative 24; Mismatches 73; Indels. 60; Gaps 11;

QY 70 IHSQNDLVEYSPTVEKHLTFDGMTV-----RELCSAALITSDNRNAILLETTI---G 118
    :| | | | | : :| | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Query Match 5.28; Score 69.5; DB 15; Length 104;
Best Local Similarity 24.78; Pred. No. 42;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9

QY      41 FPMSTFKVL-----LCGAVLSRIDAGOEOLGRRIHYSQNDLVESPV-----TEKHITDG 91
      1 | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      23 FCMSTERLLVYPYMLNGSVASRL-----RSLNGKRALDMPPRKRISLG 67
      1 | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      92 -----MIVRELCSAIIYMSDNTAANLL-----TTIGCPKEITAFLEHNGDVTFLDWM-- 140
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      68 AARGLLIYHEQDPRIIRDVKAANITLDEYFEAVVSGFGLAKLLDRSDSHYTAIVAGTV 127
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      141 ---EPLENEAIPNDERDTTPVAMATLTKRLITGEL---LTIASROQ--LIDMA 186
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      128 GHIAIEYILSTGSSSEK--TDVFGGILLLELTQKALDGRANQGVMLDWMV 179
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA

```



```
QY 55 -----VLSRIDAGOEOLGRIRHYSNDLYEX-----SPYTEKHLTDGMYRELCSAATM 104
DB 62 NIPAQIFVEFDASGIGILLKEQGHDLADYIFANTGTGSLHYFDGOSQRRVGIGTGG 121
QY 105 SDNTAANLLTTTIGCKEILAFLNMGDHYTRLDRWEPELNEAIPNDEROTTPVAMATT 164
DB 122 GMIOGIGYLLSQITDYKQLT-----DMAQGDGDR-----NTIDLKVRHIYKDEPPI----- 167
QY 165 LKRLTGELLTLASRQQLIDMEAD 189
DB 168 ----PGD-LTAAFGVHLHLDAD 186
```

RESULT 15

```
US-09-815-242-12979
; Sequence 12979, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haasbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OR INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12979
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12979
```

Query Match 4.9%; Score 65.5; DB 9; Length 189;
Best Local Similarity 20.5%; Pred. No. 1.1e+02;
Matches 42; Conservative 32; Mismatches 84; Indels 47; Gaps 8;

```
QY 4 TLVYKADADQAGAVGYEIDLSNGEILSFSEERFPMSTFKVLLCGA----- 54
DB 10 TLIKI--VQBDNQRTFTETLTKNIDQYVEMLNQO-----IKLCITGNGAGYIAENI 61
QY 55 -----VLSRIDAGOEOLGRIRHYSNDLYEX-----SPYTEKHLTDGMYRELCSAATM 104
DB 62 NIPAQIFVEFDASGIGILLKEQGHDLADYIFANVTGTSILHYFDGOSQRRVGIGTGG 121
QY 105 SDNTAANLLTTTIGCKEILAFLNMGDHYTRLDRWEPELNEAIPNDEROTTPVAMATT 164
DB 122 GMIOGIGYLLSQITDYKQLT-----DMAQGDGDR-----NTIDLKVRHIYKDEPPI----- 167
QY 165 LKRLTGELLTLASRQQLIDMEAD 189
DB 168 ----PGD-LTAAFGVHLHLDAD 186
```

Search completed: September 10, 2003, 12:33:20
Job time : 17.5714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_37S_157T
Perfect score: 1345
Sequence: 1 HETLVKRVDAEDQLEGRVCG.....TMDERNRQIAETIGASLIRKH 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	534	39.0	105	2	JC2566 bla protein - Pseu
2	77	5.7	191	2	S67447 hypothetical prote
3	76.5	5.7	200	2	G97064 spore coat protein
4	74.5	5.5	113	2	T45195 hypothetical prote
5	74.5	5.5	172	2	AD3606 molycoprotein bios
6	74.5	5.5	184	2	T21126 ADP-ribosylation f
7	74.5	5.5	195	2	T36975 hypothetical prote
8	72	5.4	149	2	F71252 nucleoside-diphosp
9	71.3	5.3	152	2	T36984 hypothetical prote
10	71	5.3	153	2	E81708 conserved hypotet
11	71	5.3	177	2	A83753 conserved hypotet
12	71	5.3	192	2	G83096 conserved hypotet
13	70	5.2	108	2	E84217 conserved hypotet
14	70	5.2	170	2	A80192 conserved hypotet
15	69.5	5.2	167	2	D87360 conserved hypotet
16	68.5	5.1	145	2	F84251 conserved hypotet
17	68.5	5.1	180	2	C71869 conserved hypotet
18	68.5	5.1	190	2	T09136 conserved hypotet
19	68.5	5.1	192	2	A83587 conserved hypotet
20	68	5.1	177	2	T48420 conserved hypotet
21	67.5	5.0	131	2	AD2281 conserved hypotet
22	66.5	4.9	116	2	C82906 conserved hypotet
23	66.5	4.9	177	2	D90227 conserved hypotet
24	66.5	4.9	196	2	H69647 conserved hypotet
25	66.5	4.9	198	2	D95285 conserved hypotet
26	66	4.9	146	2	C72703 conserved hypotet
27	65.5	4.9	42	2	F66978 collagen alpha 1(X
28	65.5	4.9	181	2	A81902 hydrogenase small
29	65.5	4.9	195	2	A60623 probable bacterioph

30	65.5	4.9	197	2	H90211 conserved hypotet
31	65	4.8	148	2	E75283 conserved hypotet
32	65	4.8	150	2	T06385 calmodulin - soybe
33	65	4.8	160	1	E69186 conserved hypotet
34	65	4.8	168	2	B75498 conserved hypotet
35	65	4.8	177	1	B43387 conserved hypotet
36	65	4.8	177	2	J01931 nonstructural prot
37	65	4.8	180	2	G70912 nonstructural prot
38	64.5	4.8	145	2	H75262 hypothetical prote
39	64.5	4.8	151	2	D81333 hypothetical prote
40	64	4.8	113	2	D70580 probable protein-t
41	64	4.8	162	2	AG0769 hypothetical prote
42	64	4.8	168	2	C85715 probable acetyltra
43	64	4.8	178	2	I40124 unknown protein en
44	64	4.8	191	2	E95333 outer surface prot
45	64	4.8	197	2	AC2133 hypothetical prote
					iron(III) dictitrat

ALIGNMENTS

RESULT 1

JC2566 bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C:Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H., Schweitzer, H.P., Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.

Gene 14811281, 81-86, 1994

A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MES>

A:Genetics:

A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match

39.0% Score 524; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 1,1e-36;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTIRKLITGELTTLASROOLIDMEADKVAGPLRSALPAGFIADKSGAGRSRG 219

DB 2 AMATTIRKLITGELTTLASROOLIDMEADKVAGPLRSALPAGFIADKSGAGRSRG 61

QY 220 ITAALGPDGKRSRIIVYITTSQATMDERNRQIAETIGASLIRKH 263

DB 62 ITAALGPDGKRSRIIVYITTSQATMDERNRQIAETIGASLIRKH 105

RESULT 2

S67447 hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

Submitted to the EMBL Data Library, March 1996

A:Reference number: Z21766

A:Accession: T38062

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:269944; NID:g1217974; PID:CAA93808.1; PID:g1217978; GSPDB:C

C:Genetics:

A:Experimental source: strain 972n-; cosmid c1F12

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match

5.7% Score 77; DB 2; Length 191;

Conserved hypothetical protein TC0378 [Imported] - Chlamydia muridarum (strain Nigg)
Species: Chlamydia muridarum Chlamydia muridarum

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: E81708
 R:Rad, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequence of *Chlamydia trachomatis* Mogen and *Chlamydia pneumoniae* AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: E81708
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-153 <STO>
 A:Cross-references: GB:AE002305; GB:AE002160; NID:g7190418; PIDN:AAF39236.1; PID:g719041
 A:Experimental source: strain N19g (MOPn)
 C:Genetics:
 A:Gene: TC0378

Query Match 5.3%; Score 71; DB 2; Length 153;
 Best Local Similarity 34.6%; Pred. No. 92;
 Matches 18; Conservative 10; Mismatches 16; Indels 8; Gaps 2;

OY 124 TATLHNMGDHVTLRDRP--ELNEALPNDEROTTTPVAMATLRKLTGEL 173
 DB 101 TTVAREGHEVTLTSEPPNTDLMQODKTPP-----RREMLSGFL 146

RESULT 11
 AB3753
 hypothetical protein BH0825 [Imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: AB3753
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: AB3650; MUID:20512582; PMID:11058132
 A:Accession: AB3753
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-177 <STO>
 A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04544.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0825
 C:Superfamily: *Bacillus subtilis* hypothetical protein ykkA

Query Match 5.3%; Score 71; DB 2; Length 177;
 Best Local Similarity 20.3%; Pred. No. 11e+02;
 Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

OY 42 PMWTFKVLGAVLRIDAGQQLGRRIHVSQNDLVEYSPYTKHLLTDGTVRELCSAA 101
 DB 11 PEMOTSVGLFYAMVEEND-----RLH-----HLIEDYTEELYYKG 47

OY 102 TMSDNTAANLLTTIGPKELTAFLLHNMGDHVTLRD--RWPEL--NEAIPND----- 151
 DB 48 SDGDENSMQQL-----NHLTVAVKWFPRIRGELIPDSLEAHGPM 89

OY 152 -BRDTPPVAMATLRKLTGELTLASRQ-----LIDNEADKVAAPLRSLAP 201
 DB 90 VDKDKLPLVVTSLVSQVELKGRYVALALKERCAALHDDDLARWLPYEERQATIRMGJ- 148

OY 202 AGWFLAD 208
 DB 149 --WHMAD 153

RESULT 12
 G83096
 conserved hypothetical protein PA4399 [Imported] - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
 C:Accession: G83096
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Ladbly, K.; L
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: AB2950; MUID:20437337; PMID:10964043
 A:Accession: G83096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <STO>
 A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA07787.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4399
 C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
 Best Local Similarity 26.0%; Pred. No. 1.3e+02;
 Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

OY 53 GAVLSRI-----DAGEQL--GRRIHVSQNDLVEYSPYTKHLLTDGTVRELCSA----- 100
 DB 2 GNLRSKTYRTGDRGETGLAGRRVPSHPRIEAGVADLNSQLGLLAELELRGAMP 61

OY 101 -----ATMSDNTAANLLTTIGPKELTAFLLHNMGDHVTLR-----DRWPEL-----N 145
 DB 62 GLEEIYQALAPVQHR-----LPDLGGLAMPETRALDETVANLSCIDRMNDELGPLKN 116

OY 146 EALPNDERDTPPVAMATLRKLTGELTLASRQQLIDMEADYVAGPLRSALPAGWF 205
 DB 117 FILPGSR-----PVAQAHYCRSLAR-----SAERQCALQDEETLDSGLVRLYLRNLSLLF 168

OY 206 IADKSGAGERSNGIT--AALGPD 227
 DB 169 VAARATARRGVAEILMEAAKPD 192

RESULT 13
 E84217
 hypothetical protein Vng0594h [Imported] - *Halobacterium* sp. NRC-1
 C:Species: *Halobacterium* sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84217
 R:Ng, W.V.; Kennedy, S.P.; Mahafiras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky
 ; Leitnuser, B.; Keller, K.; Cruz, R.; Dawson, M.J.; Hough, D.W.; Maddocks, D.C.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
 A>Title: Genome sequence of *Halobacterium* species NRC-1.
 A:Reference number: AB4160; MUID:20504483; PMID:11016950
 A:Accession: E84217
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA019105.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG0594H

Query Match 5.2%; Score 70; DB 2; Length 108;
 Best Local Similarity 27.5%; Pred. No. 69;
 Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

OY 44 MSTFKVLGAVLRIDAGQQLGRRIHVSQNDLVEYSPYTKHLLTDGTVRELCSAAT 103
 DB 1 MDTTVTELGLALLALENTD-----RVFEVRFDALEVTDTLRRHGDV-----GSIV 50

OY 104 MSDNTAANLLTTIGPREL-----TATLHNMGDHVTLRDR 139
 DB 51 NDDGTRTMARLTVPQDSDFIAVEVPTSFVAIVDAATRTDR 92

RESULT 14
 AB0192
 conserved hypothetical protein YP01575 [Imported] - *Yersinia pestis* (strain CO92)

[illegible]

Search completed: September 10, 2003, 12:26:19
Job time : 11.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds
(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E_37S_157T
Perfect score: 1345
Sequence: 1 HPEITLVKVKDAEDQIGARVGV.....TMDERNRQIAIGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	5.7	191	1 YD4A_SCHPO	Q10346 schizosacch
2	75	5.6	201	1 RACG_DICDI	Q9G9S0 dictyosteli
3	72	5.4	149	1 NDR_TREPA	Q83974 treponema p
4	68.5	5.1	196	1 RAC2_LOTUA	Q40320 lotus japon
5	68	5.1	122	1 HC3L_THIPE	P80509 thlobacilli
6	66.5	4.9	116	1 RBR4_TREPA	Q9P9H0 ureaplasma
7	66.5	4.9	196	1 ALKH_BACSU	P50846 b khg/kdpg
8	65.5	4.9	206	1 KTHY_METAC	Q8THS9 methanosarc
9	65	4.8	177	1 VNSC_RINDK	P35948 rinderpest
10	65	4.8	177	1 VNSC_RINDK	P35948 rinderpest
11	63.5	4.7	144	1 MARR_ECOLI	P27245 escherichia
12	63.5	4.7	185	1 YCIC_ECOLI	P38522 escherichia
13	63	4.7	176	1 HSLV_THEMA	Q9HYZ1 thermotoga
14	62.5	4.6	146	1 HBC_RABIT	P02099 oryctolagus
15	62.5	4.6	150	1 PDUV_SALTY	Q94DM6 salinimella
16	62.5	4.6	160	1 TATB_HELPY	Q94DM6 helicobacte
17	62.5	4.6	172	1 YDEJ_ECOLI	P31131 escherichia
18	62.5	4.6	182	1 PYRE_STPCO	Q9487 streptomyce
19	62.5	4.6	184	1 HRP1_PSEST	P37929 pseudomonas
20	62.5	4.6	195	1 TRPF_THEVO	Q979V6 thermoplasma
21	62.5	4.6	200	1 RR4_PELNE	Q9FSD9 pellia nees
22	62	4.6	173	1 Y265_BORBU	O51280 borrelia bu
23	62	4.6	174	1 IHBB_RAT	P17491 rattus norv
24	62	4.6	182	1 RR4_BELCH	O19990 belamanda
25	62	4.6	182	1 YR6L_BIFLO	Q8G520 bifidobacte
26	62	4.6	186	1 RRP_RICPR	Q94E08 rickettsia
27	62	4.6	186	1 YCEB_SALTY	P40822 salinimella
28	61.5	4.6	152	1 YU33_YERPE	O8ACF8 yersinia pe
29	61.5	4.6	178	1 HSLV_RALSO	O8ACF8 ralslonia s
30	61	4.5	121	1 SECR_HUMAN	P06683 homo sapien
31	61	4.5	184	1 ARU2_DROME	O08849 drosophila
32	61	4.5	194	1 RR4_TRIEN	O20234 itis ensata
33	61	4.5	200	1 TATB_CAUDR	O9A6T1 caulobacter

ALIGNMENTS

RESULT 1	YD4A_SCHPO	STANDARD:	PRT:	191 AA.
ID	YD4A_SCHPO			
AC	010346:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Hypothetical protein ClF12.04c in chromosome I.			
GN	SPAC1F12.04C			
OS	Schizosaccharomyces pombe (Fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RM	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE-21848401; PubMed-11859360;			
RA	Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holtroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,			
RA	James K., Jones L., Jones M., Leathers S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quayl M.A., Rabinowitsch E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellton J., Slimmons M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volkart G., Aert R., Robben J., Grymptre B.,			
RA	Wetjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Filiz C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Wambitt R., Punnelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaude V., Mottier S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,			
RA	Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Domínguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,			
RA	Shpakowski G.V., Ussery D., Barrett B.G., Nurse P.,			
PT	"The genome sequence of Schizosaccharomyces pombe."			
RL	Nature 415:871-880(2002).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; 269944; CAA93808.1; -			
DR	PIR; T38062; S67447			
DR	GenDB_Spombe; SPAC1F12.04c; -			

34	60.5	4.5	103	1 RS10_NEIGO	P48851 neisseria g
35	60.5	4.5	156	1 BFR_AZOVI	P22759 azotobacter
36	60.5	4.5	160	1 TATB_HELPY	Q94M58 helicobacte
37	60.5	4.5	184	1 MLR1_SCHPO	Q94M58 schizosacch
38	60	4.5	148	1 CALM_BLAEM	O94FV6 blastoclad
39	60	4.5	178	1 PYRE_ARCFU	O28533 archaeglob
40	60	4.5	178	1 UCRL_ANASP	P70758 anabaena sp
41	59.5	4.4	103	1 RS10_NEIMA	O57312 neisseria m
42	59.5	4.4	135	1 CCRN_PAROL	O57312 paracanthu
43	59.5	4.4	184	1 RL6_MYCPN	O50303 mycoplasma
44	59.5	4.4	196	1 I196_ASFB7	P27943 african svl
45	59.5	4.4	198	1 OGGL_ARCFU	O29876 archaeglob

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ribosome-binding factor A.
 CM RBRF OR 00321.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
 NCBI_TaxID=134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sever 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Caswell G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum."
 RL Nature 407:752-762(2000).
 CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
 CC with 30S subunits that are part of 70S ribosomes or polysomes).
 CC Essential for efficient processing of 16S rRNA. May interact with
 CC the 5' terminal helix region of 16S rRNA (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO THE RBRF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF002129; AAF30730.1; -
 DR HAMAP: MF_00003; - 1.
 DR InterPro: IPR000238; R1b_dind_facta.
 DR Pfam: PF02033; RBRF.1.
 DR ProDom: PD007327; R1b_dind_facta; 1.
 DR TIGRFAMs: TIGR00082; rbrf.1.
 DR PROSITE: PS01319; RBRF. FALSE_NEG.
 KW rRNA processing; complete proteome.
 KM SEQUENCE 116 AA; 1324 MW; A79DCCT1F0547514 CRC64;
 SQ
 Query Match 4.9%; Score 66.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No.1e+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;
 QY 144 LNAATPDERDPTTPVAMATTLKTLTGL-----LTASRQOLIDWNE-ADKVAQPL 196
 DB 18 INNALNEINDKIAKLARVAVR--LSNDLSVAKFLDAHKRESMLKLVENNYSG-LL 74
 QY 197 RSALPAGW-----FIADKS 210
 DB 75 RSKLAEMTSYKVPBLRFVIDER 97
 RESULT 7
 ALKH_BACSU STANDARD; PRT; 196 AA.
 AC P50846;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE KMG/KDPG aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase
 DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KMG-aldolase); 2-
 DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
 DE dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate
 DE aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
 DE aldolase)].
 DE KDCG.
 GN Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CC NCBI_TaxID=1423;
 OX [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 RA Seror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the *serA* and *kds* loci cloned in a yeast artificial chromosome."
 RL Microbiology 142:2005-2016(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berrtero M.G., Bessieres P., Boloitin A., Borchert S.,
 RA Boriss R., Bourret M., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denilcot F., Devine K.M., Ducleroit A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Filiz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Hentut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic C.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetle D., Portolillo S., Prescott A.M.,
 RA Preeanan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroter R., Shin B.S., Soldo B.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconelli E., Takagi T., Takahashi H., Takemori K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viart A., Wandut R., Wedler E., Weller H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate - pyruvate +
 CC glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-glucuronate 6-phosphate -
 CC pyruvate + D-glucuronide 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GLYOXYLATE.
 CC -1- SUBUNIT: Homotrimer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KMG/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L47838; AAB38480.1; -
 DR EMBL: Z99115; CAB14127.1; -
 DR PIR: H69647; H69647.
 DR Subtilist: BG11386; kdgA.
 DR InterPro: IPR000887; Aldolase_KDPG_KMG.
 DR Pfam: PF01081; Aldolase.1.
 DR TIGRFAMs: TIGR01182; eda; 1.
 DR PROSITE: PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
 DR PROSITE: PS00160; ALDOLASE_KDPG_KMG_2; 1.
 KW Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
 KM ACT_SITE 43 43 BY SIMILARITY.

Db 145 MWLAN 150

RESULT 10

ID	VNSC_RINDR	STANDARD	PRT	177 AA
AC	003339			
DT	01-JUN-1994	(Rel. 29, Created)		
DT	01-JUN-1994	(Rel. 29, Last sequence update)		
DT	01-FEB-1995	(Rel. 31, Last annotation update)		
DE	Nonstructural protein C.			
GN	C.			
OS	Rinderpest virus (strain RBOK) (RDV).			
OC	Vitruvius, ssRNA negative-strand viruses; Mononegavirales;			
CC	Paramyxoviridae; Paramyxovirinae; Morbillivirus.			
OX	NCBI_TaxID=36409;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-93155558; PubMed-8429304;			
RA	Barton M.D., Shalla M.S., Barrett T.;			
RT	"Cloning and sequence analysis of the phosphoprotein gene of			
RL	Rinderpest virus.";			
RL	J. Gen. Virol. 74:299-304(1993).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL; X68311; CAA48391.1;			
DR	EMBL; Z30697; CAA83179.1;			
DR	PIR; J01931; J01931.			
DR	InterPro; IPR003875; Paramyxovir_NSC.			
DR	Pfam; PF02725; Paramyxov_NSC; 1.			
DR	Nonstructural protein.			
DR	SEQUENCE 177 AA; 19926 MW; 76D8D46A6D3FB07 CRC64;			
QY	Query Match	4.8%; Score 65; DB 1; Length 177;		
QY	Best Local Similarity	24.6%; Pred. NO. 2.2e+02;		
QY	Matches	31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;		

Db 145 MWLAN 150

RESULT 11

ID	MARR_ECOLI	STANDARD	PRT	144 AA
AC	P27245; P76882; P77582;			
DT	01-AUG-1992	(Rel. 23, Created)		
DT	15-OCT-2001	(Rel. 40, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Multiple antibiotic resistance protein marr.			
GN	MARR OR SOXO OR CPXB OR INAR OR B1530.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			

RX MEDLINE-93186717; PubMed-8383113;

RA Cohen S.P., Haechler H., Levy S.B.;

RT "Genetic and functional analysis of the multiple antibiotic

RL resistance (mar) locus in Escherichia coli.";

RL J. Bacteriol. 175:1484-1492(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE-97426517; PubMed-9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1234-1238(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE-97251357; PubMed-9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,

RA Sakemoto G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,

RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RL corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

RN [4]

RP CHARACTERIZATION.

RX MEDLINE-94110219; PubMed-8282690;

RA Ariza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;

RT "Repressor mutations in the marRAB operon that activate oxidative

RT stress genes and multiple antibiotic resistance in Escherichia

RL coli.";

RL J. Bacteriol. 176:143-148(1994).

RN [5]

RP CHARACTERIZATION.

RX MEDLINE-95286534; PubMed-7768850;

RA Seoane A.S., Levy S.B.;

RT "Characterization of MarR, the repressor of the multiple antibiotic

RT resistance (mar) operon in Escherichia coli.";

RL J. Bacteriol. 177:3414-3419(1995).

RN [6]

RP MOTAGENESIS.

RX MEDLINE-20223625; PubMed-10760140;

RA Aleksun M.N., Kim Y.S., Levy S.B.;

RT "Mutational analysis of MarR, the negative regulator of marRAB

RT expression in Escherichia coli, suggests the presence of two regions

RT required for DNA binding.";

RL Mol. Microbiol. 35:1394-1404(2000).

CC -1- FUNCTION: REPRESSOR OF THE MARRAB OPERON WHICH IS INVOLVED IN THE

CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS

CC GENES. BINDS TO THE MARO OPERATOR/PROMOTER SITE.

CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; M96235; AAC16394.1; ALT_INIT.

DR EMBL; AE000250; AAC74603.1; ALT_INIT.

DR EMBL; D90795; BAA15212.1;

DR EMBL; D80796; BAA15220.1;

DR EMBL; D90797; BAA15232.1;

DR PIR; E64907; E64907.

DR PIR; JGGS; 28-DEC-01.

DR Ecogene; EG11435; marr.

```

DR InterPro: IPR000835; HTH_MARR.
DR Pfam: PF01047; MARR. 1.
DR PRINTS: PR00598; HTHMARR.
DR SMART: SM00347; HTH_MARR. 1.
DR PROSITE: PS01117; HTH_MARR. 1.
KW Transcription regulation; DNA-binding; 1.
KW Antibiotic resistance; Complete proteome; 3D structure.
FT MARGEN 45 45 V->E: INCREASED TRANSCRIPTION OF THE
FT MARGEN 77 77 REGION II TRANSCRIPT.
FT MARGEN 123 144 REGION II TRANSCRIPT.
FT MARGEN 144 AA: 16065 MW; BE7DE549E24DD1 CRC64;
SEQUENCE MISSING: INCREASED TRANSCRIPTION OF THE
REGION II TRANSCRIPT.
Query Match 4.7%; Score 63.5; DB 1; Length 144;
Best Local Similarity 29.5%; Pred. No. 2.3e+02;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;
QY 66 LGRRIHY--SQND--LYEV--SP--VTEKHITDGMVRELCSAATMSDNTAANLLTTIG 118
DB 14 LGLLHMVWOKKORLNEYISPLDITPAQKFKVLSIR--CAACIT----- 56
QY 119 GPELTAFLH-NMGDVTFLDR-----WEPELNEAIPNDRD-----TTTPVAMATTIR 167
DB 57 -PVELKKVLSVDIGALTRMLDRVCKGWERLPN--PNDKRGVYKLTGGALICECHQ 113
QY 168 LITGELLTLASROQLIDMMEADKVA--GPLLRSALP 201
DB 114 LVGQDL-----HQELTKNLTADEVATLEVLKKRVL 144
RESULT 12
YCJC_ECOLI STANDARD: PRT; 185 AA.
AC P38522; P76839; P77417;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ycjC.
GN YCJC OR B1299 OR SF1304.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-E.COLI: STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT *The complete genome sequence of Escherichia coli K-12.*;
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-E.COLI: STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alida H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashiwano K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Samei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RT *A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.*;
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 65-185 FROM N.A.
RC SPECIES-E.COLI:
RX MEDLINE-91216440; PubMed-1840553;

```

```

RA Helm R., Strehler E.E.;
RT *Cloning an Escherichia coli gene encoding a protein remarkably
RT similar to mammalian aldehyde dehydrogenases.*;
RL Gene 99:15-23(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-S.flexneri: STRAIN-301 / Serotype 2a;
RX MEDLINE-2272406; PubMed-12384590;
RA Jin O., Yun Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT *Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.*;
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [5]
RP IDENTIFICATION.
RC SPECIES-E.COLI:
RA Rudd K.E.;
RT Unpublished observations (AUG-1994).
RN [5]
CC -1- SIMILARITY: SOME, TO H.INFLUENZAE H10659.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC CC
CC EMBL; AE000228; AAC74381.1; -.
CC EMBL; D90768; BAA14868.1; -.
CC EMBL; D90767; BAA14859.1; -.
CC EMBL; M38433; -. NOT ANNOTATED_CDS.
CC EMBL; AE015157; AAN42915.1; -.
CC PIR; F64878; F64878.
CC ECoGene; EG12431; YCJC.
CC InterPro: IPR007113; CUPIn_sup.
CC InterPro: IPR001387; HTH_3.
CC Pfam: PF01381; HTH_3; 1.
CC SMART; SM00530; HTH_XRE_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 185 AA; 20091 MW; D7D0D3FD794A7768 CRC64;
Query Match 4.7%; Score 63.5; DB 1; Length 185;
Best Local Similarity 27.5%; Pred. No. 3.1e+02;
Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;
QY 145 NEAIRNDERDPTTPVAMATTIRKLLTTCGLTLASROQLIDMMEADKVAAGPLRSALPAGW 204
DB 34 HSAISTIQDKVSPA--ISTLQKLKLVGSLSE-----FSPSEPKPDEPV----- 78
QY 205 FIADKSGAGSGSGRSIIAALGPDGKPSRIIVY-----YTTGSAATWDER-NQAIETIGASL 259
DB 79 -VINODLIEGSGVGSVKLVHNGNPNTLAFETTOPGT--TTGERIKHGEIEGYVL 135
RESULT 13
HSIV_THEMA STANDARD: PRT; 176 AA.
ID HSIV_THEMA
AC Q9WY21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease hslV (EC 3.4.25.-).
GN HSIV OR TM0521
OS Thermotoga maritima
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;

```

```

RX MEDLINE-99287316; PubMed-10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.W.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eilen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermococcus maritima.
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PERTIDASE FAMILY T1B. HSIV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001728; AAD35606.1; -.
CC PIR; G72365; G72365.
CC HSSP; P31059; 1E94.
CC MEROPS; T01.006; -.
CC TIGR; TM0521; -.
CC HAMAP; MF_00248; -.
CC InterPro; IPR001353; Proteasome-protease.
CC Pfam; PF00227; Proteasome; 1.
CC HydroLase; Protease; Complete proteome.
CC ACT SITE 6 By similarity.
CC SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 3.2e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 IGPEELTAFILHNGDHYTRLDRWPELENAIPDERRTTPVAMATTLRKLGLGELTL 176
DB 42 LSEGKVLAFGAFGADVADATLTFDREAKLR-----WGKLLTK 78
QY 177 ASROQLIDMEADKVGKPLRSALPAGWFIADK-----SGAGERSGKII-----AAL 224
DB 79 AAVELAKWM-RTDRY---LR-RLEALLLVADKEKIFISGNGE-----VIQPPDDAAAI 127

QY 225 GPDC 228
DB 128 GSGG 131

RESULT 14
HGB_RABIT STANDARD: PRT; 146 AA.
AC P02099;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin gamma chain (Beta-3).
GN HGB.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-82053017; PubMed-6271761;
RA Hardison R.C.;
RT "The nucleotide sequence of rabbit embryonic globin gene beta 3."
RL J. Biol. Chem. 256:11780-11786(1981).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-89178632; PubMed-2486295;
RA Margot J.B., Demers G.W., Hardison R.C.;
RT "Complete nucleotide sequence of the rabbit beta-like globin gene
RT cluster. Analysis of intergenic sequences and comparison with the
RT human beta-like globin gene cluster."
RL J. Mol. Biol. 205:15-40(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE
CC GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE
CC HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN.
CC -1- SUBUNIT: Heterotetramer of two alpha chains and two gamma chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M18818; AAA02984.1; -.
CC EMBL; V00883; CAA24252.1; -.
CC PIR; A02417; HBRB3.
CC HSSP; P02100; 1A9W.
CC InterPro; IPR002337; Beta_haem.
CC InterPro; IPR000971; Globin.
CC Pfam; PR00042; globin; 1.
CC PRINTS; PR00814; BETAHAEM.
CC PROSITE; PS01035; GLOBIN; 1.
CC Heme; Oxygen transport; Transport; Erythrocyte; Embryo.
CC INT. MET 0
CC SEQUENCE 146 AA; 16093 MW; EB8D6C1C24DD2D82 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 146;
Best Local Similarity 30.8%; Pred. No. 2.8e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

QY 35 FRSEERPMSTFEVILGCAVLSRIDAGQDGLRRRIHYSQNDLVEYSPYTEKHLTDGMTY 94
DB 3 FTAEEKAIITSTWK-----LVYEDDAGALRL-----LVYI-PWTQRFDSFQNL 48
QY 95 RELCSAATWSNDTANLLTTGGPKELTAFILHNGDHYTRL 138
DB 49 SS-SSAIMGNPVKYKAH-----GKKVLTAF-----GDAVKND 79

RESULT 15
PDUV_SALTY STANDARD: PRT; 150 AA.
AC Q9YDM6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Propanediol utilization protein pduV.
GN pduV OR STM2056.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2;
RX MEDLINE-99429843; PubMed-10498708;
RA Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
RT "The propanediol utilization (pdu) operon of Salmonella enterica
RT serovar typhimurium LT2 includes genes necessary for formation of
RT polyphedral organelles involved in coenzyme B(12)-dependent 1, 2-
RT propanediol degradation."
RL J. Bacteriol. 181:5967-5975(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-LTF2 / SGSC1412 / ATCC 700720;
RX MEDLINE-21534948; Pubmed-11677609;
RA McClelland M., Sanderson K.E., Spleth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LTF2*;
RL Nature 413:852-856(2001).
CC -I- INDUCTION: By propenadiol.
CC -I- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF026270; AAD39020.1; -
DR EMBL: AE008790; AAL20960.1; -
DR ScyGene: SG10680; pduV.
KW ATP-binding; Complete proteome.
FT NP_BIND 8 15 ATP (POTENTIAL)
SQ SEQUENCE 150 AA; 16348 MW; 7771229432F97E56 CRC64;

```

Query Match

Best local similarity 4.6%; Score 62.5; DB 1; Length 150;

Matches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

```

QY 52 CGAVLSRIDAGEQGRRIHYSNDLYEVSPTERKHLTDGNTVREICSAITWSDNTAAN 111
DB 12 CG-----KTSLSQSLKCEALHKTKQAIEWSPMAID--TPGEYLENRCIYSALITSACEAD 65
QY 112 LLTTTGGPRELTAFLHNNGDVTRLDREWEPEINEAIPNDRDRTTPVAMATTIRRLNG 171
DB 66 VIALVYNADAWNSPF-----SPGFAPMNRPTIGLVTKADLAEFORISIVA 111
QY 172 ELLTTLASRQQLDMEADKVVAGPLRSALPA 202
DB 112 EMLTQAGAGQIF-----ITSALNNSGIDA 135

```

Search completed: September 10, 2003, 12:20:52
Job time : 6.57143 secs

THIS PAGE BLANK (USPTO)
THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345

Sequence: 1 HPELVKVKAEQDLGARVG.....TMDERNQIAIGASLIKHW 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_23:*
2: SP_Archea:*
3: SP_Bacteria:*
4: SP_Fungi:*
5: SP_Human:*
6: SP_Invertebrate:*
7: SP_Mammal:*
8: SP_mhc:*
9: SP_Organelle:*
10: SP_Phage:*
11: SP_Plant:*
12: SP_Rodent:*
13: SP_Virus:*
14: SP_Vertebrate:*
15: SP_Unclassified:*
16: SP_Virus:*
17: SP_Bacteriophage:*
18: SP_Archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	524	39.0	105	2	052026 pseudomonas
2	524	39.0	105	2	052330 pseudomonas
3	499	37.1	145	2	08BMD8 pseudomonas
4	488	36.3	102	2	052639 pseudomonas
5	464	34.5	138	2	099QF3 pseudomonas
6	464	34.5	138	2	099QF3 pseudomonas
7	464	34.5	139	2	099QF3 pseudomonas
8	462	34.3	138	2	099QF3 pseudomonas
9	462	34.3	138	2	099QF3 pseudomonas
10	462	34.3	139	2	099QF3 pseudomonas
11	459	34.1	139	2	099QF3 pseudomonas
12	338	25.1	67	2	053553 shigella fl
13	336	24.2	95	2	08GDES klebsiella
14	310	22.4	62	2	08GDES klebsiella
15	301	22.4	128	2	08VQ00 shigella fl
16	251	18.7	180	2	08KVT2 staphylococ

17	241	17.9	48	2	09RIH0 proteus mir
18	238	17.7	52	2	09RA12 shigella fl
19	197	14.6	38	2	097145 escherichia
20	169.5	12.6	134	2	08VUL3 staphylococ
21	124.5	9.3	109	2	053698 staphylococ
22	105	7.8	109	2	053698 staphylococ
23	87	6.5	100	2	093S05 escherichia
24	82	6.1	68	2	093S05 escherichia
25	80.5	6.0	202	16	098520 rhizobium l
26	80	5.9	198	2	09ACM8 streptococ
27	76.5	5.7	200	16	097JEB clostridium
28	75	5.6	181	16	08D108 yersinia pe
29	74.5	5.5	113	2	049970 mycobacteri
30	74.5	5.5	112	16	08YBW3 mycobacteri
31	74.5	5.5	172	16	08YBW3 mycobacteri
32	74.5	5.5	184	5	045379 caenorhabd
33	74.5	5.5	195	16	09RIAO streptomyc
34	74	5.5	153	5	08SS20 clostrid
35	73	5.4	131	2	09XSH0 yersinia en
36	73	5.4	175	17	08ETB8 yersinia en
37	73	5.4	201	17	08ETB8 yersinia en
38	72.5	5.4	145	2	005984 methanococ
39	72	5.4	205	6	09N275 staphylococ
40	72	5.4	205	16	098J18 staphylococ
41	71.5	5.3	152	16	09RI91 rhizobium l
42	71.5	5.3	196	2	09AH34 pseudomonas
43	71	5.3	150	10	0941G4 pseudomonas
44	71	5.3	153	16	09PKT4 nicotiana t
45	71	5.3	177	16	09KEM7 bacillus m

ALIGNMENTS

RESULT 1

ID 052026 PRELIMINARY; PRT; 105 AA.
AC 052026;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN B.la.
OS Pseudomonas aeruginosa.
OG Plasmid PR01614.
OC Bacteria: Proteobacteria: Gammaproteobacteria: Pseudomonadales;
OC Pseudomonadaceae: Pseudomonas.
OX NCBI_TaxID=287;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;
RT *Construction of improved Escherichia-Pseudomonas shuttle vectors
RT derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.*;
RL Gene 148:81-86 (1994).
DR EMBL: L30112; AAA6058.1; -;
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR pfam: PF00144; Beta_lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER
SQ SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.6e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTTLKLTGELTFLASROQLDMEADRVAGPILRSALPGWFTADSKSGAGSGSRG 219
DB 2 AMATTTLKLTGELTFLASROQLDMEADRVAGPILRSALPGWFTADSKSGAGSGSRG 61

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 62 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 105

RESULT 2
ID 052330 PRELIMINARY; PRT; 105 AA.
AC 052330;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Tnl Dia protein (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
RX MEDLINE-90264294; PubMed-2160936;
RA Kornacki J.A., Burlage R.S., Figurski D.H.;
RT "The k11-kor regulon of broad host-range Plasmid RK2: Nucleotide
RT sequence, polypeptide product and expression of regulatory gene
RT korc".
RL J. Bacteriol. 172:3040-3050(1990).
DR EMBL: M32794; AAA26408.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
SQ SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2,6e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 219
|||||
Db 2 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 61

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 62 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 105

RESULT 3
ID 08RTD8 PRELIMINARY; PRT; 145 AA.
AC 08RTD8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE SHV-5 enzyme (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
NCBI_TaxID=573;
RX MEDLINE-90264294; PubMed-2160936;
RA Kornacki J.A., Burlage R.S., Figurski D.H.;
RT "The k11-kor regulon of broad host-range Plasmid RK2: Nucleotide
RT sequence, polypeptide product and expression of regulatory gene
RT korc".
RL J. Bacteriol. 172:3040-3050(1990).
DR EMBL: M32794; AAA26408.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
SQ SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2,6e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 219
|||||
Db 2 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 61

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 62 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 105

RESULT 4
ID 052639 PRELIMINARY; PRT; 102 AA.
AC 052639;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN BLA.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=287;
RX MEDLINE-94336757; PubMed-8058819;
RA Jandons I., Touchle G., Sharp R., Almqvist K., Farinha M.A., Lam J.S.,
RT "Deletion and transposon mutagenesis and sequence analysis of the por
RT P01600 Orlr region found in the broad-host-range plasmids of the por
RT series".
RL Plasmid 31:265-274(1994).
DR EMBL: L22691; AAA98312.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 3e-34;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 219
|||||
Db 2 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 61

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 62 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 105

RESULT 5
ID 099GF3 PRELIMINARY; PRT; 138 AA.
AC 099GF3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 37.1%; Score 499; DB 2; Length 145;
Best Local Similarity 68.8%; Pred. No. 5,8e-35;
Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

OY 100 AATMSDNTAANILTTTIGPKELTAFLNMGDHYTRLDWPELNEAIPNDERDTTPV 159
|||
Db 7 AATMSDNTAANILTTTIGPKELTAFLNMGDHYTRLDWPELNEAIPNDERDTTPV 66

OY 160 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 219
|||||
Db 67 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 126

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 127 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 144

RESULT 4
ID 052639 PRELIMINARY; PRT; 102 AA.
AC 052639;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN BLA.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=287;
RX MEDLINE-94336757; PubMed-8058819;
RA Jandons I., Touchle G., Sharp R., Almqvist K., Farinha M.A., Lam J.S.,
RT "Deletion and transposon mutagenesis and sequence analysis of the por
RT P01600 Orlr region found in the broad-host-range plasmids of the por
RT series".
RL Plasmid 31:265-274(1994).
DR EMBL: L22691; AAA98312.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 3e-34;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 219
|||||
Db 2 AMATTLRKLTLGELLTLASROOLIDWMDKAVAGPLRLSALPAGWFIADKSGAGERSRG 61

OY 220 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 263
|||||
Db 62 IIAALGPDGKPSRIVVITTTGSOATMDERNROIAEIGASLIRKM 105

RESULT 5
ID 099GF3 PRELIMINARY; PRT; 138 AA.
AC 099GF3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
OC Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327049; AK07464.1; -
DR EMBL: AF327051; AK07466.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 5.3e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFILHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAAATTLRLKLT 170
DB 10 NLLATVGGPAGLTAFILQIGDVTYRLDRWETELNEALPDANDTTTPASMAATLRLKLT 69
QY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADSGAGERSGRTIAGLPGDKP 230
DB 70 SQRLSARSQRLQMWVDVAGPLIRSVLPAGWFIADTGTGASKRGARGIVALGPNKKA 129
QY 231 SRIIVYIY 237
DB 130 ERIIVYIY 136
Db 130 ERIIVYIY 136

RESULT 6
Q9AMA1 PRELIMINARY; PRT; 138 AA.
ID Q9AMA1;
AC Q9AMA1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1B/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327050; AK07465.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 5.3e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFILHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAAATTLRLKLT 170
DB 10 NLLATVGGPAGLTAFILQIGDVTYRLDRWETELNEALPDANDTTTPASMAATLRLKLT 69
QY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADSGAGERSGRTIAGLPGDKP 230
DB 70 SQRLSARSQRLQMWVDVAGPLIRSVLPAGWFIADTGTGASKRGARGIVALGPNKKA 129
QY 231 SRIIVYIY 237
DB 130 ERIIVYIY 136
Db 130 ERIIVYIY 136

RESULT 7
Q9AMA2 PRELIMINARY; PRT; 139 AA.
ID Q9AMA2;
AC Q9AMA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327048; AK07463.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 139;
Best Local Similarity 68.5%; Pred. No. 5.4e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFILHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAAATTLRLKLT 170
DB 10 NLLATVGGPAGLTAFILQIGDVTYRLDRWETELNEALPDANDTTTPASMAATLRLKLT 69
QY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADSGAGERSGRTIAGLPGDKP 230
DB 70 SQRLSARSQRLQMWVDVAGPLIRSVLPAGWFIADTGTGASKRGARGIVALGPNKKA 129
QY 231 SRIIVYIY 237
DB 130 ERIIVYIY 136
Db 130 ERIIVYIY 136

RESULT 8
Q9AMA0 PRELIMINARY; PRT; 138 AA.
ID Q9AMA0;
AC Q9AMA0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RC STAIN-E/98 9-1; Navaratnam P.;
RA Subramaniam G.; Navaratnam P.;
RT "SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
isolates from UMMC, Malaysia";
RL Submitted (DRC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327052; AK07467.1; -.
DR HSSP; P14557; ISHV.
DR InterPro; IPR001466; Beta_lactamase.
DR InterPro; IPR000871; Beta_lactamase_A.
DR Pfam; PF00144; beta-lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1 138
FT NON_TER 1 138
SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Qy	231	SRI	VIX	237
Db	130	ERIV	LY	136

Q8GDE5

192 AC 193

—

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds

(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_37S

Sequence: 1345
1 HPEITLVKVADEQICARVG.....TMDERNRQIAETGASLIRKM 263

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	583	43.3	196	7 AAP60627	Beta-urogastrone -
2	567	42.2	146	22 AAU23221	Novel human enzyme
3	567	42.2	146	22 AAU23221	Novel human enzyme
4	488	38.3	127	18 AAW20440	Human immune/haema
5	473	35.2	129	22 AAU23220	H. pylori cytoplas
6	473	35.2	129	22 AAU23220	Novel human enzyme
7	473	35.2	182	22 AAG27917	Human immune/haema
8	463	34.4	94	21 AAB59052	Novel human diagno
9	400	29.7	159	7 AAP60628	Breast and ovarian Beta-urogastrone -

10	372	27.7	101	22 AAG27935	Novel human diagno
11	329	24.5	88	22 AAG27919	Novel human diagno
12	195	14.5	119	7 AAP60626	Beta-urogastrone -
13	145	10.8	51	21 AAY92783	Veggs-beta-lactama
14	85	6.3	200	18 AAW55524	H. pylori ORF 29ep
15	79.5	5.9	158	22 AAU45138	Protonibacterium
16	78	5.8	202	24 AAE32763	Human zcyto24 prot
17	78	5.8	202	24 AAO16283	Mouse IMX129840-2
18	71	5.3	159	19 AAM69170	N-terminally tagge
19	70	5.2	15	10 AAP98503	Sequence encoded b
20	69.5	5.2	170	22 AAU56330	Protonibacterium
21	69	5.1	134	21 AAG27220	zee mays protein f
22	69	5.1	178	22 AAG91926	c glutamicum prote
23	69	5.1	180	21 AAG20206	Arbidopsis thalia
24	69	5.1	194	22 AAG81764	S. epidermidis ope
25	67.5	5.0	184	21 AAB23446	Pinus radiata cell
26	67.5	5.0	187	23 AAU97104	Human MK61 protein
27	67	5.0	101	22 AAU66278	Protonibacterium
28	67	5.0	168	22 ABB68717	Drosophila melanog
29	67	5.0	202	22 AAG82686	S. epidermidis ope
30	67	5.0	202	24 AAE32764	Human zcyto25 prot
31	66.5	4.9	177	21 AAB41582	Human ORFX ORF1346
32	66.5	4.9	196	21 AAY73344	HPRM clone 0258181
33	66	4.9	170	22 AAU52472	Protonibacterium
34	65.5	4.9	50	17 AAU03591	Human alpha 2 C4 a
35	65.5	4.9	136	21 AAY44229	Partial corn extra
36	65.5	4.9	152	22 AAG19138	Novel human diagno
37	65.5	4.9	152	22 AAG27771	Novel human diagno
38	65	4.8	179	22 AAU87627	Novel signal trans
39	65	4.8	179	22 AAU17192	Novel central nerv
40	65	4.8	179	22 AAG74084	Human colon cancer
41	65	4.8	183	19 AAG85905	S. pneumoniae deri
42	65	4.8	185	22 AAU87344	Novel central nerv
43	65	4.8	203	18 AAB31266	Neospora caninum a
44	64.5	4.8	102	23 ABB34479	Human dehydrogena
45	64.5	4.8	161	18 AAY11282	Streptococcus pneu

ALIGNMENTS

RESULT 1	
AAU60627	AAU60627 standard; Protein; 196 AA.
ID	AAU60627 standard; Protein; 196 AA.
XX	
AC	AAU60627;
XX	
DT	25-MAR-2003 (updated)
DT	17-JUN-1991 (first entry)
XX	
DE	Beta-urogastrone - beta-lactamase fusion protein from pUC2101.
XX	
KW	Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
KW	fusion protein; beta-lactamase.
XX	
OS	Synthetic.
XX	
FX	
FX	Key
FX	Protein
FX	Misc-difference
FX	Protein
FX	Label= adaptor
FX	Label= beta-urogastrone
FX	Label= beta-urogastrone
FX	DE3523634-A.
FX	09-JAN-1986.
FX	02-JUL-1985;
FX	02-JUL-1984;
FX	84JP-0137691.

PA (EART) EARTH CHEM CO LTD.
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX WPI: 1986-015031/03.
DR N-PSDB: AAN60631.
XX
FT New gene for expression of beta-urogastrone - its derive., plasmd(s)
PT and transformed cells confg. It.
XX
PS Disclosure; Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 196 AA;
Query Match 43.3%; Score 583; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. No. 6,86-52;
Matches 115; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 HPELVKVKDAEDQIGARVGYEILDNSGKILSFSEHRRPMMSTFKVLGAVLSRID 60
DB 24 HPELVKVKDAEDQIGARVGYEILDNSGKILSFSEHRRPMMSTFKVLGAVLSRID 83
OY 61 AGQEDLGRHRYSONDLVEYSPVTEKHLTDGTVRELCSAITYSDNTAANLLTTI 117
DB 84 AGQEDLGRHRYSONDLVEYSPVTEKHLTDGTVRELCSAITYSDNTAANLLTTI 140
RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209667.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

CC	diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders and infectious disorders (e.g. reproductive disorders e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy.
CC	AAU22915-AAU23814 represent the novel human enzyme polypeptides of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
CC	
XX	Sequence 146 AA:
SQ	
Query Match	42.2%; Score 567; DB 22; Length 146;
Best Local Similarity	97.4%; Pred. No. 2e-50;
Matches 112; Conservative	0; Mismatches 3; Indels 0; Gaps 0
OY	1 HPELVKVKDDEDDGARGVYEIDLNSGKIIEFSRSEPRFPMSTFYVLGCAVLSRID 60 HPELVKVKDDEDDGARGVYEIDLNSGKIIEFSRSEPRFPMSTFYVLGCAVLSRID 87
Db	
OY	61 AGOEDLGRIHIYSQNDLVEYSPVTEKHLDGKTVELCSAATNMSDNTAANLLT 115 88 AGOEGLGRIRIHSQNDLVEYSPVTEKHLDGKTVELCSAATNMSDNTAANLLT 142
Db	
RESULT 3	
AAAG90870	
ID	AAAG90870 standard; Protein: 146 AA.
AC	
XX	AAAG90870;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen SEQ ID NO:18463.
XX	
KW	Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX	
KM	Cytostatic; gene therapy; vaccine; metastasis.
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
PD	
XX	
PD	09-AUG-2001.
XX	
Pf	17-JAN-2001; 2001MO-US01354.
XX	
XX	31-JAN-2000; 2000US-0179065. PR 04-FEB-2000; 2000US-0180628. PR 24-FEB-2000; 2000US-0184664. PR 02-MAR-2000; 2000US-0186350. PR 16-MAR-2000; 2000US-0189874. PR 17-MAR-2000; 2000US-0190076. PR 18-APR-2000; 2000US-0198123. PR 19-MAY-2000; 2000US-0205515. PR 07-JUN-2000; 2000US-0209467. PR 28-JUN-2000; 2000US-0214886. PR 30-JUN-2000; 2000US-0215135. PR 07-JUL-2000; 2000US-0216647. PR 07-JUL-2000; 2000US-0216880. PR 11-JUL-2000; 2000US-0217487. PR 11-JUL-2000; 2000US-0217496. PR 14-JUL-2000; 2000US-0218280. PR 26-JUL-2000; 2000US-0220963. PR 26-JUL-2000; 2000US-0220964. PR 14-AUG-2000; 2000US-0224518. PR 14-AUG-2000; 2000US-0224519. PR 14-AUG-2000; 2000US-0225213. PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0235266.
 PR 14-AUG-2000; 2000US-0235267.
 PR 14-AUG-2000; 2000US-0235268.
 PR 14-AUG-2000; 2000US-0235270.
 PR 14-AUG-2000; 2000US-0235447.
 PR 14-AUG-2000; 2000US-0235757.
 PR 14-AUG-2000; 2000US-0235758.
 PR 14-AUG-2000; 2000US-0235759.
 PR 18-AUG-2000; 2000US-0236279.
 PR 22-AUG-2000; 2000US-0236681.
 PR 22-AUG-2000; 2000US-0236868.
 PR 23-AUG-2000; 2000US-0237182.
 PR 30-AUG-2000; 2000US-0238924.
 PR 01-SEP-2000; 2000US-0239289.
 PR 01-SEP-2000; 2000US-0239343.
 PR 01-SEP-2000; 2000US-0239344.
 PR 01-SEP-2000; 2000US-0239345.
 PR 05-SEP-2000; 2000US-0239509.
 PR 05-SEP-2000; 2000US-0239511.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233397.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0235935.
 PR 13-OCT-2000; 2000US-0235937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 01-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 BA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483426/52.
 XX N-PSDB; AAK63651.
 DR
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 XX
 PS Claim 11; SEQ ID NO 18463; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX

SO Sequence 146 AA;
Query Match 42.2%; Score 567; DB 22; Length 146;
Best Local Similarity 97.4%; Pred. No. 2e-50;
Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 HPETLVKVAEDQGLGARVGYIELDLSGKILSFSEERFPMMSTFKVLLCGAVLSRID 60
Db 28 HPETLVKVAEDQGLGARVGYIELDLSGKILSFSEERFPMMSTFKVLLCGAVLSRID 87
OY 61 AGGEOGLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCAAITMSDNTANLLT 115
Db 88 AGGEOGLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCAAITMSDNTANLLT 142
RESULT 4
AAW20440 standard; protein; 127 AA.
AC AAW20440;
XX
DT 14-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 34574062.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
KW replication; transcription; recombination; repair.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 121
FT Misc-difference 122 /note="encoded by codon YGC"
FT /note="encoded by codon WGG"
XX
XX W09640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgaard BL;
XX
DR WPI; 1997-052306/05.
XX
DR N-PSDB; AAT67613.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 61; Pages 615-616; 1481pp; English.
XX
CC The present sequence is a Helicobacter pylori cytoplasmic protein
CC involved in genomic replication, transcription, recombination and repair.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SO Sequence 127 AA;
Query Match 36.3%; Score 488; DB 18; Length 127;
Best Local Similarity 99.0%; Pred. No. 2.4e-42;
Matches 96; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 HPETLVKVAEDQGLGARVGYIELDLSGKILSFSEERFPMMSTFKVLLCGAVLSRID 60
Db 24 HPETLVKVAEDQGLGARVGYIELDLSGKILSFSEERFPMMSTFKVLLCGAVLSRID 83
OY 61 AGGEOGLGRRIHYSQNDLVEYSPVTEKHLTDGMTVREL 97
Db 84 AGGEOGLGRRIHYSQNDLVEYSPVTEKHLTDGMTVREL 120
RESULT 5
AAU23220 standard; Protein; 129 AA.
ID AAU23220
XX
AC AAU23220;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #306.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
XX
XX W020015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.

PR	16-AUG-2000	2000US-0226379
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229387
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231988
PR	14-SEP-2000	2000US-0232357
PR	14-SEP-2000	2000US-0232358
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234984
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239337
PR	13-OCT-2000	2000US-0239338
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0244674
PR	08-NOV-2000	2000US-0244675
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246538
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611

Query Match	35 2%	Score 473:	DB 22:	Length 129:
Best Local Similarity	96.9%	Pred. No. 8.9e-41:		
Matches 93:	Conservative 0:	Mismatches 3:	Indels 0:	Gaps 0:

```
OY      1 HPELVYKDAEDDGLGARVGIETIEDLNSKILFSEFSEERPPMNSTFKVLGCAVLSRID 60
Db      28 HPELVYKDAEDDGLGARVGIETIEDLNSKILFSEFSEERPPMNSTFKVLGCAVLSRID 87
QY      61 AGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMYRE 96
Db      88 AGQEQLGRRIHYSQNDLVEYSPVTEKHLTDGMYRXE 123

RESULT 6
AAM90871
ID      AAM90871 standard; Protein; 129 AA.
XX
AC      AAM90871;
XX
DE      07-NOV-2001 (first entry)
XX
XX      Human Immune/haematopoietic antigen SEQ ID NO:18464.
KW      Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW      cytostatic; gene therapy; vaccine; metastasis.
OS      Homo sapiens.
XX
PN      WC200157182-A2.
XX
PD      09-AUG-2001.
XX
PF      17-JAN-2001; 2001MO-US01354.
XX
PR      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0186874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205515.
PR      07-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      14-AUG-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225265.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226281.
PR      22-AUG-2000; 2000US-0226686.
PR      22-AUG-2000; 2000US-0227183.
PR      23-AUG-2000; 2000US-0227189.
PR      30-AUG-2000; 2000US-0228924.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      05-SEP-2000; 2000US-0229509.
PR      05-SEP-2000; 2000US-0229513.

PR      06-SEP-2000; 2000US-0230437.
PR      06-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232397.
PR      14-SEP-2000; 2000US-0232398.
PR      14-SEP-2000; 2000US-0232399.
PR      14-SEP-2000; 2000US-0232400.
PR      14-SEP-2000; 2000US-0232401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      14-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234224.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0235368.
PR      25-SEP-2000; 2000US-0235369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237038.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237040.
PR      13-OCT-2000; 2000US-0239335.
PR      20-OCT-2000; 2000US-0240960.
PR      20-OCT-2000; 2000US-0241221.
PR      20-OCT-2000; 2000US-0241785.
PR      20-OCT-2000; 2000US-0241786.
PR      20-OCT-2000; 2000US-0241787.
PR      20-OCT-2000; 2000US-0241808.
PR      20-OCT-2000; 2000US-0241809.
PR      20-OCT-2000; 2000US-0241826.
PR      01-NOV-2000; 2000US-0244617.
PR      08-NOV-2000; 2000US-0246474.
PR      08-NOV-2000; 2000US-0246475.
PR      08-NOV-2000; 2000US-0246476.
PR      08-NOV-2000; 2000US-0246477.
PR      08-NOV-2000; 2000US-0246478.
PR      08-NOV-2000; 2000US-0246523.
PR      08-NOV-2000; 2000US-0246524.
PR      08-NOV-2000; 2000US-0246525.
PR      08-NOV-2000; 2000US-0246526.
PR      08-NOV-2000; 2000US-0246527.
PR      08-NOV-2000; 2000US-0246528.
PR      08-NOV-2000; 2000US-0246532.
PR      08-NOV-2000; 2000US-0246609.
PR      08-NOV-2000; 2000US-0246610.
PR      08-NOV-2000; 2000US-0246611.
PR      08-NOV-2000; 2000US-0246613.
PR      17-NOV-2000; 2000US-0249207.
PR      17-NOV-2000; 2000US-0249208.
PR      17-NOV-2000; 2000US-0249209.
PR      17-NOV-2000; 2000US-0249210.
PR      17-NOV-2000; 2000US-0249211.
PR      17-NOV-2000; 2000US-0249212.
PR      17-NOV-2000; 2000US-0249213.
PR      17-NOV-2000; 2000US-0249214.
PR      17-NOV-2000; 2000US-0249215.
PR      17-NOV-2000; 2000US-0249216.
PR      17-NOV-2000; 2000US-0249217.
```

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249329.
PR 17-NOV-2000; 2000US-0249330.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0239678.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
DR N-PSDB; AAK63652.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX Claim 11: SEQ ID NO 18464; 3071pp + Sequence Listing: English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 129 AA:
SQ
Query Match 35.2%; Score 473; DB 22; Length 129;
Best Local Similarity 96.9%; Pred. No. 8.9e-41;
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 HPEITLVKKAEDDQAGRVGTEIDLNSGKILFSRSEPRPMSTFVLLCGAVLSRID 60
DB 28 HPEITLVKKAEDDQAGRVGTEIDLNSGKILFSRSEPRPMSTFVLLCGAVLSRID 87
OY 61 AGQEQIGRRIRHYSQNDLVEYSPTVEKHLTDGKATVE 96
DB 88 AGQEQIGRRIRHYSQNDLVEYSPTVEKHLTDGKATVE 123
RESULT 7
ABG27917
ID ABG27917 standard; Protein; 182 AA.
XX
XX ABG27917;

XX 18-FEB-2002 (first entry)
DF Novel human diagnostic protein #27908.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PE 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSB-) HYSB INC.
PA Drmanac RT, Liu C, Tang YF;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS92104.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20: SEQ ID NO 58276; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 182 AA:
SQ
Query Match 35.2%; Score 473; DB 22; Length 182;
Best Local Similarity 87.3%; Pred. No. 1.5e-40;
Matches 96; Conservative 3; Mismatches 7; Indels 4; Gaps 1;
OY 1 HPEITLVKKAEDDQAGRVGTEIDLNSGKILFSRSEPRPMSTFVLLCGAVLSRID 60
DB 24 HPEITLVKKAEDDQAGRVGTEIDLNSGKILFSRSEPRPMSTFVLLCGAVLSRID 83
OY 61 AGQEQIGRRIRHYSQNDLVEYSPTVEKHLTDGKATVE 110
DB 84 AGQEQIGRRIRHYSQNDLVEYSPTVEKHLTDGKATVE 129
RESULT 8
AAB59052
ID AAB59052 standard; Protein; 94 AA.
XX
XX AAB59052

XX AAB59052;
AC

XX 27-MAR-2001 (first entry)
DT

XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
DE

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antifungal; antiparasitic and cardiant activity. The
KW particularly breast and ovarian cancer. The nucleic acid sequences,
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.
KW

OS Homo sapiens.

XX WO200055173-A1.
PN

XX 21-SEP-2000.
PD

XX 08-MAR-2000; 2000WO-US05861.
PF

XX 12-MAR-1999; 99US-0124270.
PR

XX (HUMA-) HUMAN GENOME SCI INC.
PA

XX Rosen CA, Ruben SM;
PI

XX WPI; 2000-611515/58.
XX

XX N-PSDB; AAF21955.
DR

XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX

XX Claim 11; Page 1228; 1299pp; English.
PS

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antitumor; antitumor; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiant activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemia; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX

XX Sequence 94 AA;

Query Match 34.4%; Score 463; DB 21; Length 94;
Best Local Similarity 98.9%; Pred. No. 6.1e-40;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 75 NDLYEYSPYTEKHLTDGTVRELCSAATMSDNTAANLLTTTGPKRL7FLNMGDHY 134
DB 1 NDLYEYSPYTEKHLTDGTVRELCSAATMSDNTAANLLTTTGPKRL7FLNMGDHY 60
OY 135 TRLDWPEPLNEAIPNDERDTPMPVAAATT 164
DB 61 TRLDWPEPLNEAIPNDERDTPMPVAAATT 90

RESULT 9

XX AAB60628
ID AAB60628 standard; Protein; 159 AA.
XX

XX AAB60628;
AC

XX 25-MAR-2003 (updated)
DT

XX 17-JUN-1991 (first entry)
DE

XX Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
DE

XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
KW fusion protein; beta-lactamase.
KW

XX Synthetic.
OS

XX Key Location/Qualifiers
FH 1..96
FT /label= beta-lactamase

XX Misc-difference 97.100
FT /label= adaptor

XX Protein 101..159
FT /label= beta-urogastrone

XX DE3523634-A.
PN

XX 09-JAN-1986.
PD

XX 02-JUL-1985; 85DE-3523634.
PF

XX 02-JUL-1984; 84JP-0137691.
PR

XX (EART) EARTH CHEM CO LTD.
PA

XX Aoki S, Ohgal H, Horinaka A, Hiramatsu H, Kounoto S, Nishimura A;
PI Matsushiro S;
PI

XX WPI; 1986-015031/03.
DR

XX N-PSDB; AAN60632.
DR

XX New gene for expression of beta-urogastrone - its derive., plasmid(s)
PT and transformed cells contg. it.
PT

XX Disclosure; Page 59-61; 92pp; German.
PS

XX The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628 and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX

XX Sequence 159 AA;

Query Match 29.7%; Score 400; DB 7; Length 159;
Best Local Similarity 68.3%; Pred. No. 4.3e-33;
Matches 86; Conservative 8; Mismatches 20; Indels 12; Gaps 2;

OY 1 HPEVLVVKDAEDQAGRVGYIELDINSKILSFSEERFPKMSFFKVLGCAVYSRD 60
DB 24 HPEVLVVKDAEDQAGRVGYIELDINSKILSFSEERFPKMSFFKVLGCAVYSRD 83
OY 61 AGOEOGRRIHYSONDLYERS-----PYTEKHLTDGTVRELCSAATMSDNTAANL 112
DB 84 AGOEOGRRIHYSONDLYERSAKRNSDSECPLSH----DGYCLHDGVCMTIEALDKVACNC 139
OY 113 LATTIG 118
DB 11

Db 140 VGCIG 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein: 101 AA.
AC ABG27935;
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #27926.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR N-PSDB: AAS92122.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PT
XX
XX Claim 20; SEQ ID No 58294; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 101 AA;

Query Match 27.7%; Score 372; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.7e-30;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 DTTMPVAMATTLLKRLTGLTGLTASRQQLDWMKADKVGAPLLRSALPGWFIADKSGAG 213
DB 10 DTTMPVAMATTLLKRLTGLTGLTASRQQLDWMKADKVGAPLLRSALPGWFIADKSGAG 69

OY 214 ERGSRGIIAALGPD 227
DB 70 ERGSRGIIAALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein: 88 AA.
AC ABG27919;
XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #27910.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR N-PSDB: AAS92106.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PT
XX
XX Claim 20; SEQ ID No 58278; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 88 AA;

Query Match 24.5%; Score 329; DB 22; Length 88;
Best Local Similarity 81.2%; Pred. No. 4e-26;
Matches 69; Conservative 2; Mismatches 4; Indels 10; Gaps 1;

OY 4 TLVKKADKEDDGLGAVGYTIELDLSGKLTIESFRSEERPPMNSTPKVLLCGAVLSRIDAQ 63
DB 4 TLVKKADKEDDGLGAVGYTIELDLSGKLTIESFRSEERPPMNSTPKVLLCGAVLSRIDAQ 63

DB 1 TLVKKVDAEDQAGARVGYIELDLNSGKILESFRPERFPMSTFKVLLCGAVLSRIDAGQ 60
 QY 64 EQLGRRI-----HYSQNDLY 78
 |||||
 DB 61 EQLGRRIILRWTLSTHQSOKSL 85

RESULT 12

AAPE0626
 ID AAP0626 standard; Protein; 119 AA.
 XX
 AC AAP0626;
 XX

25-MAR-2003 (updated)
 DT 17-JUN-1991 (first entry)
 XX

Beta-urogastrone - beta-lactamase fusion protein from pUG2301.
 DE XX

Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 KW fusion protein; beta-lactamase.
 XX

OS Synthetic.
 XX

Key Location/Qualifiers
 FH Protein 1..63
 FT /label= beta-lactamase

Misc-difference 64..66
 FT /label= adaptor
 FT 67..119

Protein /label= beta-urogastrone
 FT
 XX

DE3523634-A.
 PN
 XX

09-JAN-1986.
 PD
 XX

02-JUL-1985; 85DE-3523634.
 PR
 XX

02-JUL-1984; 84JP-0137691.
 PR
 XX

(EART) EARTH CHEM CO LTD.
 PA
 XX

Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
 PI Matsushiro S;
 XX

WPI; 1986-015031/03.
 DR
 XX

N-PSDB; AAN60630.
 DR
 XX

New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 PT and transformed cells contg. it.
 PT
 XX

Disclousure; Page 55; 92pp; German.
 PS
 XX

The fusion protein is less easily degraded by proteases and so
 CC protects beta-urogastrone and beta-lactamase collects in the periplasm
 CC of E.coli. It is therefore easy to collect and purify the product.
 CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 CC stomach acid secretion and promotes cell growth, so is useful for
 CC treating ulcers and wounds. Previously the product was obtl. only
 CC in small amts. from human urine.
 CC See also AAN60628, and 30-32.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 XX

Sequence 119 AA:
 SQ

Query Match 14.5%; Score 195; DB 7; Length 119;
 Best Local Similarity 97.5%; Pred. No. 4.5e-12;
 Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HETLVKVDADQAGARVGYIELDLNSGKILESFRSEER 40
 |||||
 DB 24 HETLVKVDADQAGARVGYIELDLNSGKILESFRSEER 63

RESULT 13

AAV92783
 ID AAV92783 standard; Protein; 51 AA.
 XX
 AC AAV92783;
 XX

29-AUG-2000 (first entry)
 DT
 XX

Vtgs-beta-lactamase fusion protein (partial).
 DE XX

Vtgs; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation;
 KW beta-lactamase.
 XX

Chimeric - Oreochromis aureus.
 OS Chimeric - Synthetic.
 OS

Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Vtgs

Cleavage-site 15..16
 FT Protein 22..51
 FT /note= "beta-lactamase mature protein"

W0200026366-A1.
 PN
 XX

11-MAY-2000.
 PD
 XX

29-OCT-1999; 99MO-SG00108.
 PR
 XX

30-OCT-1998; 98US-0106426.
 PR
 XX

26-OCT-1999; 99US-0426776.
 PR
 XX

(UYST-) UNIT SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 PA

Ding JL, Tan NS, Ho B;
 PI
 XX

WPI; 2000-365615/31.
 DR
 XX

N-PSDB; AAA28507.
 DR
 XX

Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX

Example 6; Fig 14A; 73pp; English.
 PS

A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated PBADVtglactkana was
 CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 CC
 XX

Sequence 51 AA:
 SQ

Query Match 10.8%; Score 145; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 2e-07;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PETLVKVDADQAGARVGYIELDLNSGK 30
 |||||
 DB 23 PETLVKVDADQAGARVGYIELDLNSGK 51

```

RESULT 14
AAM55524
ID AAM55524 standard; Protein: 200 AA.
XX
AC AAM55524;
XX
DT 02-JUL-1998 (first entry)
XX
DE H. pylori ORF 29ep10720_24432762_c3_39 cellular protein.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
OS Helicobacter pylori.
XX
PN M09737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTRA ) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI: 1997-503122/46.
DR N-PSDB: AAV24933.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Clalms 14,93; Page 726-727; 1145pp; English.
XX
XX This sequence is a H. pylori cellular protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence
CC of H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 200 AA:
XX
Query Match 6.3%; Score 85; DB 18; Length 200;
Best Local Similarity 26.5%; Pred. No. 2.2;
Matches 52; Conservative 31; Mismatches 43; Indels 70; Gaps 13;
OY 22 IELDINSGRILESF---RSEERFPMASTFKVLLCGAVTSRDAGEQLGRIRHSQNDLV 78
DB 11 VVLDPDSQSKMETFAIRAEKERPFISLF-----NRSSGFSDTLKQWVSKENYLLI 61
OY 79 -----EISPTTEKHLVDGMTVRELCSAATMSDNTAANLL--TTIGCKREL-TATLHMNG 131

```

```

DB 62 DFKGEYSKETOK-----AMLS-----NIVLVPTT--PSQDTEVIAMNL 99
OY 132 DHVTRLDKRWEPPELNE---AIPNDEKDTTPVAMATTIKRLTGLTGLTASRQLIDNNE- 187
DB 100 ERIEQILO-----EINEMLRALIVINRWPTIP-----TIKERQALIEFIKE 139
OY 188 ---ADKVGAPLLRSAL 200
DB 140 NNPSDKRIT--LLESSL 153

RESULT 15
AAU45138
ID AAU45138 standard; Protein: 158 AA.
XX
AC AAU45138;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #6034.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosia; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN W0200181581-A2.
XX
PD 01-NOV-2001.
XX
PE 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208847P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Sreiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
DR N-PSDB: AAS59525.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID NO 6333; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosia and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

```

XX	Sequence	158 AA;
SQ		

Query Match	5.9%	Score 79.5;	DB 22;	Length 158;
Best Local Similarity	33.3%;	Pred. No. 5.9;		
Matches	26;	Conservative	9;	Mismatches 36;
				Indels 7;
				Gaps 2;

QY 110 ANLLTIGGPELTAIFHNMGDHTLRDREPELEINAFNDEQDTTPVAMATTLR-K 167
 Db 80 ARATSTGAPKYSSASLTARQASPTWVGPEPESTRPI-----TSAPLAFEDTLRSGR 134
 QY 168 LLTGELTLASRQQLIDW 185
 Db 135 GRMGSWVLATRIPIPLRW 152

Search completed: September 10, 2003, 12:24:51
Job time : 33.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds

(without alignments)
938.485 Million cell updates/sec

Title: SE02_37S
Perfect score: 1345

Sequence: 1 HPEITLVKVADEQAGARVG.....TMDERNROIAETGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	5.5	169	4	US-09-328-352-4491 Sequence 4491, Ap
2	72.5	5.4	124	4	US-08-311-731A-202 Sequence 202, App
3	71	5.3	139	3	US-08-991-890-4 Sequence 4, Appl1
4	70.5	5.2	197	4	US-09-252-991A-24241 Sequence 24241, A
5	70.5	5.2	198	3	US-09-413-814-87 Sequence 87, Appl1
6	65.5	4.9	158	3	US-09-010-809-19 Sequence 19, Appl1
7	65.5	4.9	178	4	US-09-252-991A-29942 Sequence 29942, A
8	64.5	4.8	203	4	US-09-252-991A-24921 Sequence 24921, A
9	64	4.8	150	3	US-09-239-909-2 Sequence 2, Appl1
10	64	4.8	203	4	US-08-624-677A-2 Sequence 2, Appl1
11	63.5	4.7	160	4	US-09-252-991A-24737 Sequence 24737, A
12	62.5	4.6	144	2	US-08-225-480-4 Sequence 4, Appl1
13	62.5	4.6	144	4	US-09-118-445-4 Sequence 4, Appl1
14	62	4.6	132	4	US-09-252-991A-22681 Sequence 22681, A
15	61.5	4.6	142	4	US-09-345-473E-31 Sequence 31, Appl1
16	61.5	4.6	170	4	US-08-858-207A-519 Sequence 519, App
17	61	4.5	174	2	US-08-557-122A-6 Sequence 6, Appl1
18	61	4.5	174	4	US-09-263-666-6 Sequence 6, Appl1
19	61	4.5	199	4	US-09-252-991A-30363 Sequence 30363, A
20	61	4.5	200	2	US-08-557-122A-12 Sequence 12, Appl1
21	61	4.5	200	4	US-09-267-666-12 Sequence 12, Appl1
22	61	4.5	167	2	US-08-690-849-2 Sequence 2, Appl1
23	60.5	4.5	167	2	US-09-004-053-2 Sequence 2, Appl1
24	60.5	4.5	145	4	US-09-134-001C-4982 Sequence 4982, Ap
25	60	4.5	184	4	US-09-795-926-46 Sequence 46, Appl1
26	60	4.5	184	4	US-09-732-210-1282 Sequence 1282, Ap
27	59.5	4.4	103	4	US-09-732-210-1282 Sequence 1282, Ap

28	59.5	4.4	146	4	US-09-134-001C-5269 Sequence 5269, Ap
29	59.5	4.4	171	4	US-09-252-991A-22351 Sequence 22351, A
30	59	4.4	136	4	US-09-252-991A-30622 Sequence 30622, A
31	59	4.4	148	4	US-09-134-001C-3640 Sequence 3640, Ap
32	59	4.4	171	4	US-09-107-532A-3979 Sequence 3979, Ap
33	59	4.4	180	4	US-09-194-905-5 Sequence 5, Appl1
34	59	4.4	189	4	US-08-671-548C-48 Sequence 48, Appl1
35	58.5	4.3	102	3	US-08-479-744A-47 Sequence 47, Appl1
36	58.5	4.3	102	3	US-08-280-757B-47 Sequence 47, Appl1
37	58.5	4.3	174	4	US-09-072-596-281 Sequence 281, App
38	58	4.3	168	2	US-08-451-947-6 Sequence 6, Appl1
39	58	4.3	168	2	US-08-424-826A-6 Sequence 6, Appl1
40	58	4.3	168	2	US-08-928-69A-6 Sequence 6, Appl1
41	58	4.3	168	4	US-08-450-842-6 Sequence 6, Appl1
42	58	4.3	168	4	US-08-451-390-6 Sequence 6, Appl1
43	58	4.3	168	5	PCR-0591-06950-6 Sequence 6, Appl1
44	58	4.3	183	4	US-09-252-991A-31347 Sequence 31347, A
45	58	4.3	187	4	US-09-252-991A-29325 Sequence 29325, A

ALIGNMENTS

```
RESULT 1
US-09-328-352-4491
; Sequence 4491, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4491
; LENGTH: 169
; TYPE: PRY
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4491
```

Query Match
Best Local Similarity 23.3%; Pred. No. 3.3;
Matches 37; Conservative 34; Mismatches 62; Indels 26; Gaps 7;

```
OY 118 GGPETAFLEHMGDHYRLDRMEPELNEALPN--DESDTTPYAMATTLKRLTG---E 172
|| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 GGENTIMKMLAQDN---RFDWVEQIVQYLDRLTVRERIMVFTTIFVVIVGYSIMK 61
OY 173 LTLTASHQQ-----LIDMMEADKYAGPLRSALPGWFIADKSGAGER--GSGIIT 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 MHSIAEQOQRKLNLDKDLVWVWQSNV-----TMKRNANLEIDKSKRIORVAAQOGILV 115
OY 223 ALGPDGPRSRIVYITTGSOATMDERNROIAETGASLIK 261
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 116 SSQNGEQDLQIV--THQNTAIIANFLTOLAQGLSTIOK 152
```

```
RESULT 2
US-08-311-731A-202
; Sequence 202, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
```

```

STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-2441
TELEFAX: 617/720-3500
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202

Query Match          5 4%; Score 72.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. NO. 3;
Matches   34; Conservative 18; Mismatches    44; Indels   23; Gaps     8;

OY      130 GDHY--TRLDREBELN---EALPNDERDT----TFVNAVATLRLKLTGELLTLASRQ 180
        ||| : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | |
DB      12 MGDDIGMRBERMGNTQCPLRVYVGDSPTLDGRASPEDILIT--NLSPTIMSHPPS 69
        ||| : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | |

OY      181 QLIIDMEA-DKYAGPLL-----NSALPAWFIADKSAGSGRGSGITIALGPDGKPSR 232
        ||| : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | |
DB      70 RDDDWVEFDPLDQAVPDATGDKATMPAVGI---GASTFGS-GILASISPFROPAR 123
        ||| : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | |

RESULT 3
US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6114307
GENERAL INFORMATION:
APPLICANT: Jaspers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
TITLE OF INVENTION: CONKLIN, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:

```

[illegible]


```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24921
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24921

Query Match
Best Local Similarity 27.1%; Score 64.5; DB 4; Length 203;
Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7;

QY 135 TRLDWREPEL-----NEAIPNDERDTPVAMATTLRLKLTGELLTLASROQLDPMDE 187
DB 34 TPLDSSSRRLRVSLAKRKAMPRAQARPMKLAET--RALTGLSTLTSSR-----DMS- 86
QY 188 ADKVAGPLRSALPAGFIADKSGSGRSGRNGIIALGPDGKPSRIVITYTGSQATMDE 247
DB 87 -----TSAMPGGGL-----SPGWLASTRLBCK-AKLAIFTATT--AAKLV 123
QY 248 RNRQIAEIG 256
DB 124 PTRYSTIG 132

RESULT 9
US-09-239-909-2
; Sequence 2, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kumbo Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve M
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: EP 99300136.1
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

Query Match
Best Local Similarity 4.8%; Score 64; DB 3; Length 150;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SONDLVEYSPYTEKHLTDG---MTVRELCSAATMSDNTAANLLTTT-----GGPRE 122
DB 6 SEEQIVDFKEANGLDQDGCITVEELATVIRSLDQPTREELQDMISEVDADONGTIE 65
QY 123 LTAFLHNGDHTRLDRKEPELEAIPNDERDTPVAMATTLRLKLT--TSELTLASRO 180
DB 66 FDEFLSLAKKVKYDIDA--EELKEAFKVDKONGYIS-ASELRHVMINIGELKLTDEVE 123
QY 181 QLIDMMEAD 189
DB 124 QMI--KEAD 130

RESULT 10
US-08-624-677A-2
; Sequence 2, Application US/08624677A
```

```
; Patent No. 6476192
; GENERAL INFORMATION:
; APPLICANT: Lally, Nicola C.
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Dubey, Jitender P.
; TITLE OF INVENTION: Antigens Useful for the Serodiagnosis of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Graeter, Janelle S.
; STREET: Rm. 411, Bldg. 005, BANC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/624,677A
; FILING DATE: 15-Apr-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0228.95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-504-6629
; TELEFAX: 301-504-5060
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-624-677A-2

Query Match
Best Local Similarity 4.8%; Score 64; DB 4; Length 203;
Matches 39; Conservative 23; Mismatches 72; Indels 30; Gaps 7;

QY 54 AVLSRIDAGBOGQGRHYSQNDLVEYSPYTEKHLTDGKTVRLCSAA--ITMSDNTAAN 111
DB 33 AGVSNYDGDGDDAGNPD-----SDVTDAITDGEKPRVSGCKPHTTKGSLIK 83
QY 112 LTLTTGKPELTAFLHNGDHY-TRLDREPELEAIPNDER-DTTPVA----- 160
DB 84 LAVPVGA---LNSYL--VADRVLPELTSAGEGTESIPCKRKVTAAGIAALVAAAAFA 138
QY 161 ---MATTLRKLTGELLTLASROQLDPMDEADKVAGPLRSALP 201
DB 139 GLGLARTFRHFVPRKSKTVASSEDALGNSSEQYVEGTIVNGSSDP 182

RESULT 11
US-09-252-991A-24737
; Sequence 24737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24737
```


LENGTH: 160
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24737

Query Match 4.7%: Score 63.5; DB 4; Length 160;
Best Local Similarity 37.5%; Pred. No. 44;
Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

QY 3 ETLVKVADADQAGRVYIELDLSKIL-----ESFRSEERPMMS 45
DB 109 EILEKVSFLKSOGAAIVHLEVDENNPKAVSFKKSFERRERFVMS 156

RESULT 12

US-08-225-480-4
Sequence 4, Application US/08225480
Patent No. 5817793
GENERAL INFORMATION:
APPLICANT: LEVY, STUART B.
TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
TITLE OF INVENTION: ASSAYS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,085
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: T0359/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-225-480-4

Query Match 4.6%: Score 62.5; DB 2; Length 144;
Best Local Similarity 29.4%; Pred. No. 48;
Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

QY 66 LGRRIHY--SOND--LVEY-SP--VTEKHLTDGWTVELCSAATMSDNTAANILLTTIG 118
DB 14 LGRLIHWNOKKRLINEYLSPLDITAAQFVLCISIR--CAACIT----- 56
QY 119 GKRELTAFLH-NMGDHYTRLDR-----WEPLNEAIPNDESDTTPVAMATTLKLLTG- 171
DB 57 -PVELAKRVLSVDLGLTRMLDRLVCKGVERLPR--PNDRKG-----VLVKLTG 104
QY 172 -----ELLTLASROOLDIMWEADKVA--GPLRSALP 201
DB 105 AAIQEQCHQVAGDGLHDELTKNLTADEAVATLEYLKKVLP 144

RESULT 13
US-09-118-445-4
Sequence 4, Application US/09118445
Patent No. 6391545
GENERAL INFORMATION:
APPLICANT: LEVY, STUART B.
TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
TITLE OF INVENTION: ASSAYS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,445
FILING DATE: 17-Jul-1998
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,480
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: T0359/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-118-445-4

Query Match 4.6%: Score 62.5; DB 4; Length 144;
Best Local Similarity 29.4%; Pred. No. 48;
Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

QY 66 LGRRIHY--SOND--LVEY-SP--VTEKHLTDGWTVELCSAATMSDNTAANILLTTIG 118
DB 14 LGRLIHWNOKKRLINEYLSPLDITAAQFVLCISIR--CAACIT----- 56
QY 119 GKRELTAFLH-NMGDHYTRLDR-----WEPLNEAIPNDESDTTPVAMATTLKLLTG- 171
DB 57 -PVELAKRVLSVDLGLTRMLDRLVCKGVERLPR--PNDRKG-----VLVKLTG 104
QY 172 -----ELLTLASROOLDIMWEADKVA--GPLRSALP 201
DB 105 AAIQEQCHQVAGDGLHDELTKNLTADEAVATLEYLKKVLP 144

RESULT 14
US-09-252-991A-22681
Sequence 22681, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196, 136
CURRENT APPLICATION NUMBER: US/09/252,991A
FILING DATE: 1999-02-18

```

? PRIOR APPLICATION NUMBER: US 60/074,768
? PRIOR FILING DATE: 1998-02-18
? PRIOR APPLICATION NUMBER: US 60/094,190
? PRIOR FILING DATE: 1998-07-27
? NUMBER OF SEQ ID NOS: 3342
? SEQ ID NO 22681
? LENGTH: 132
? TYPE: PRT
? ORGANISM: Pseudomonas aeruginosa
? US-09-252-991A-22681

```

Query Match	4.68;	Score 62;	DB 4;	Length 132;
Best Local Similarity	29.88;	Pred. NO. 48;		
Matches	28;	Conservative	8;	Mismatches 34;
				Indels 24;
				Gaps 4;

```

QY      144 LNEAIPNDERDPTMPVAMATTLRKLTGELLTLASROOLIDMEADYAGPILRSALP- 201
      | : || | | | | | | | | | | | | | | | | | | | | | | | | |
DB      34 LRGTIPN---HAGRPQOATVLPMEENGE-----RRVAFPMKSPPPR 74

```

```

QY      202 AGMETADKSGAGERGSRGIIA---ALGPDGKPSR 232
          |  ||::| | | | |
Db      75 AAPEYADRAAGPDPGTGGCCARPRRAAPAGVPAR 108

```

```

RESULT 15
US-09-345-473E-31
; Sequence 31, Application US/09345473E
; Patent No. 6558903
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin
; TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
; FILE REFERENCE: 35800/183781
; CURRENT APPLICATION NUMBER: US/09/345,473E
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 31
; LENGTH: 142
; TYPE: prt
; ORGANISM: Rattus norvegicus
US-09-345-473E-31

```

Query Match	4.68;	Score 61.5;	DB 4;	Length 142;
Best Local Similarity	41.78;	Pred. NO. 61;		
Matches 15; Conservative	5;	Mismatches 11;	Indels 5;	Gaps 1.

```
Oy      210 SGAGERSRGIALGP----DGKPSRIVITTTG 240
          ||| | | | : | | | | : | |
Db      47  SGGSGNDSGCLAYLAPELLDNDGNASKASDVYFSG 82
```

Search completed: September 10, 2003, 12:31:03
Job time : 12.8571 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_37S

Sequence: 1345
1 HPEPLVKNVDAEDQLGARVG.....TMDERNRQIAIGASLIRKM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEM_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEM_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEM_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEM_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEM_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEM_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	463	34.4	94	US-10-102-806-760	Sequence 760, App
2	78	5.8	202	US-10-127-816-9	Sequence 9, App11
3	78	5.8	202	US-10-142-717-12	Sequence 12, App1
4	69.5	5.2	149	US-10-233-926-4	Sequence 4, App1
5	69	5.1	178	US-10-156-761-8136	Sequence 8136, App
6	69	5.1	149	US-09-738-626-5680	Sequence 5680, App
7	69.5	5.1	206	US-10-156-761-13867	Sequence 13867, A
8	67.5	5.0	184	US-10-101-464A-765	Sequence 765, App1
9	67.5	5.0	187	US-09-948-018-8	Sequence 8, App11
10	67	5.0	194	US-10-156-761-8623	Sequence 8623, App
11	67	5.0	202	US-10-127-816-11	Sequence 11, App1
12	65	4.8	179	US-09-764-868-757	Sequence 757, App
13	65	4.8	179	US-10-106-698-4858	Sequence 4858, App
14	65	4.8	202	US-10-189-346-12	Sequence 12, App1
15	64.5	4.8	191	US-10-156-761-8978	Sequence 8978, App

16	64.5	4.8	193	11	US-09-805-354-8	Sequence 8, App11
17	64.5	4.8	193	13	US-10-144-259-8	Sequence 8, App11
18	63.5	4.7	189	9	US-09-815-242-5862	Sequence 5862, App
19	63.5	4.7	189	9	US-09-815-242-12979	Sequence 12979, A
20	63.5	4.7	202	15	US-10-156-761-13460	Sequence 13460, A
21	63.5	4.7	206	10	US-09-738-626-5425	Sequence 5425, App
22	62.5	4.6	144	15	US-10-131-406-4	Sequence 4, App11
23	62.5	4.6	160	12	US-09-882-227-230	Sequence 230, App
24	62.5	4.6	190	15	US-10-156-761-9507	Sequence 9507, App
25	62.5	4.6	191	15	US-09-764-870-390	Sequence 390, App
26	62.5	4.6	191	15	US-10-125-540-390	Sequence 4796, App
27	62	4.6	162	10	US-09-738-626-4796	Sequence 4796, App
28	62	4.6	174	9	US-09-864-761-35777	Sequence 35777, A
29	62	4.6	190	10	US-09-738-626-4637	Sequence 4637, App
30	61.5	4.6	88	15	US-10-100-252-6	Sequence 6, App11
31	61.5	4.6	142	10	US-09-862-027-31	Sequence 31, App1
32	61.5	4.6	187	15	US-10-156-761-12111	Sequence 12111, A
33	61.5	4.6	196	10	US-09-738-626-6624	Sequence 6624, App
34	61.5	4.6	202	15	US-10-189-346-16	Sequence 16, App1
35	61	4.5	70	9	US-09-864-761-41555	Sequence 41555, A
36	61	4.5	177	11	US-09-791-932-69	Sequence 69, App1
37	61	4.5	185	15	US-10-101-464A-616	Sequence 616, App
38	61	4.5	189	9	US-09-764-870-523	Sequence 523, App
39	61	4.5	189	15	US-10-125-540-523	Sequence 154, App
40	61	4.5	203	9	US-09-800-729-154	Sequence 265, App
41	60.5	4.5	152	15	US-10-156-761-14346	Sequence 40, App1
42	60.5	4.5	196	15	US-10-219-220-265	Sequence 46, App1
43	60	4.5	134	11	US-09-768-235B-40	
44	60	4.5	184	9	US-09-798-029-4	
45	60	4.5	184	9	US-09-795-926-46	

ALIGNMENTS

RESULT 1
US-10-102-806-760
Sequence 760, Application US/10102806
Publication No. US2003005442A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA103PCT
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 760
LENGTH: 94
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (80)
OTHER INFORMATION: (80)
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760
Xaa equals any of the naturally occurring L-amino acids

Query Match 34.4% Score 463; DB 15; Length 94;
Best Local Similarity 98.9%; Pred. No. 2.4e-40;

Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

0Y 75 NDIVESPVTEKHLTDGTVRELCSAITSDDTAANLITIGPRELTAFAHNNGDHY 134
Db 1 NDIVESPVTEKHLTDGTVRELCSAITSDDTAANLITIGPRELTAFAHNNGDHY 60

OY 135 TRLDRMEPELNEAIPNDERDTMPVAMATT 164
Db 61 TRLDRMEPELNEAIPNDERXTMPVAMATT 90

RESULT 2
US-10-127-816-9

; Sequence 9, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Kluehner, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindavogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-9

Query Match 5.8%; Score 78; DB 15; Length 202;
Best Local Similarity 21.4%; Freq. No. 6;
Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LILGAVLSRIDAGQGLGRIRHYSQND--LVEYSPYTEKHL-----TDGATVREL----- 97
Db 17 LILAAVLTIRQADPPVPRATRLPEVAKDCHIAQFKSLSPRELOAFKKAADAEIKRLLENDL 76
OY 98 -CSA-----AITMS--DNTAANLLTTTGGPRELTAFLHN 129
Db 77 RCSSHLFPRAWDLKQLQVQERPKALQAEVALTLKWEKMTDSALATIIIGOPHLTSLHSHS 136
OY 130 MGDHVT-----RLDRMEPELNEAIPNDERDTMPVAMATTLRLLTGELTFL 176
Db 137 QLOCTQLOATAEPSPSRRLSLRWLHRLQEA-QSKETPGCLEASVTSNLFRLITDLCVCV 195
OY 177 ASRQOLI 183
Db 196 ANGDOCV 202

RESULT 3
US-10-142-717-12
; Sequence 12, Application US/10142717
; Publication No. US20030104579A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Baum, Peter R.
; APPLICANT: Mosley, Bruce A.
; APPLICANT: Ketchum, Randal R.
; APPLICANT: Taylor, Scott L.
; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-142-717-12

Query Match 5.8%; Score 78; DB 15; Length 202;
Best Local Similarity 21.4%; Freq. No. 6;
Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LILGAVLSRIDAGQGLGRIRHYSQND--LVEYSPYTEKHL-----TDGATVREL----- 97
Db 17 LILAAVLTIRQADPPVPRATRLPEVAKDCHIAQFKSLSPRELOAFKKAADAEIKRLLENDL 76
OY 98 -CSA-----AITMS--DNTAANLLTTTGGPRELTAFLHN 129
Db 77 RCSSHLFPRAWDLKQLQVQERPKALQAEVALTLKWEKMTDSALATIIIGOPHLTSLHSHS 136
OY 130 MGDHVT-----RLDRMEPELNEAIPNDERDTMPVAMATTLRLLTGELTFL 176
Db 137 QLOCTQLOATAEPSPSRRLSLRWLHRLQEA-QSKETPGCLEASVTSNLFRLITDLCVCV 195
OY 177 ASRQOLI 183
Db 196 ANGDOCV 202

RESULT 4
US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US2003013382A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
; FILE REFERENCE: BBI419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
US-10-233-926-4

Query Match 5.2%; Score 69.5; DB 16; Length 149;
Best Local Similarity 20.8%; Freq. No. 30;
Matches 22; Conservative 25; Mismatches 34; Indels 25; Gaps 4;

OY 156 TMPVAMATTLEKLL-----TCELLTASRQQLIDMMEADKVAGPLRLSALPGWFTADK 209
Db 11 SLSTLSXSLPLNLAMADHAAEAPOSSOEEDMKAEAGGDDV-----EVADR 61
OY 210 SGAGRGSGRIIAALGPDGKPSRIIV-----IYTTGSQATWDEHNR 250
Db 62 GGGGGAANGGI-----PEGREIRYADGIYDLFHFHGAKSLEQAKR 102

RESULT 5
US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8136
```

```
Query Match
Best Local Similarity 5.1%; Score 69; DB 15; Length 149;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;
```

```

QY 168 LITGELLT-----ASRQQLDMEADVAGPLLSALPAGMFLADKSG----- 211
Db 50 LLDIELTTLKRLILVASVDKAKEMGIDMWEHD---PALSSHADGRELAEFRRLREEI 105
QY 212 AGERSGSRIALGPDGKPSR 232
Db 106 AGLRDAQALPSAGPAEKPER 126
```

```

RESULT 6
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO: 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680
```

```
Query Match
Best Local Similarity 5.1%; Score 69; DB 10; Length 178;
Matches 34; Conservative 19; Mismatches 65; Indels 36; Gaps 5;
```

```

QY 113 LITIGCKRELTAFLNMGCH--VTRLDREPELNEAIPNDEDTTPVAMATTLKRLT 170
Db 27 LKVTKAGPKSAVAIAIKVDSRPLDQLEVSQIEGELFDAAEQRGELNGAGVTLLEVSIP 86
QY 171 GELLTLASRQQLIDMEADVAGPLLSALPAGMFLADKSG----- 230
Db 87 G-----VDNPL---TLPKRW-----RRNRRLVALDQDGRK 114
QY 231 --SRIYVITTSQATMDERNRQIAEIGASLIRH 262
Db 115 RVARIGALMDAETHVVLIERKKLLEVTLELAH 148
```

```

RESULT 7
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 13867
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867
```

```
Query Match
Best Local Similarity 5.1%; Score 69; DB 15; Length 206;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;
```

```

QY 70 IHYSQNDLVESPYTEKHLTDGMTV-----RELCSAATMSDNTAANLLTTI---G 118
Db 1 VHASQG---AAGPGTESGEEKVTMDAAGSFRFVA-----NRSSALLKTAVALISG 50
QY 119 CPKE-----LTAFLNMGDHTVRID-----RW-----EPELNEAIPNDE 152
Db 51 GDRHAEDLLQNALIKADRWRSRIDEPEAVYROYLYRQOVSRWMLKRRRRLSVAEP-PE 109
QY 153 RDTTPVAMATTLKRLILGELLTLASRQQL-----DMEDAKVAGPLLSALPAGMFL 205
Db 110 ASTGPDASAAELRLVMGALARLARQRTVLVRYREDLPEAD--VARIIGCSGVTVRS 167
QY 206 IADSGAGERSGSRIALGPDGKPSR 232
Db 168 TTHRSIARLRIAPELALGPADAEQEPSR 197
```

```

RESULT 8
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
```

```

; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-765

Query Match          5.0%; Score 67.5; DB 15; Length 184;
Best Local Similarity 24.7%; Pred. No. 65;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;

OY 41 FPMSTFKVL-----LCGAVLSRIDAGOEOLGRHHSQNDLVESPV---TEKHLTDG 91
DB 23 FCMSTERTLVYPYMLNGSVASRL-----RDSINGKPALDMPTRKRSLSG 67
OY 92 -----MTVRELCSAATMTSDNTAANLLT---TTGGPKELTAFLHNGDHYTRLDNR-- 140
DB 68 AARGLLTYHEQDDPKIIRDKAANILDEFEAVVGDGFLAKLLDHDSHTTAHVAGTV 127
OY 141 --EPBLNEAIPNDERDRTMPVAMATTIRKLLTGEL---TLASRQO--LIDRM 186
DB 128 GHIAPEVLTGSSSEK--TDVFGFGLLELTGOKALDFGRANQKVMADWV 179

RESULT 9
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: Theell et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-018-8

Query Match          5.0%; Score 67.5; DB 10; Length 187;
Best Local Similarity 22.4%; Pred. No. 66;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

OY 112 LTTTIGCKRELTAEFLHNGDHYTRLDNRPE-----LANEAIPINDERDRTMPVAMAT- 163
DB 11 LLLIALAPPEAS-----GYCGHLEWNPNNKCCSSCLQRFGP-----PPCPALERTG 58
OY 164 -----TLRKLITGELLTASR--OOLIDMMA-----DKVAGP-----L 195
DB 59 DTWKEASLPLLSRELSSLASQPLSRLLDELVELELYLADPEPGPGGMAGTTRHIA 118
OY 196 LNSALDPAGFIADKSGAGSRGIIAAGPGKPSR--IYIYTTGSOATDERRROIA 253
DB 119 ARTGLPAAV-----STFAYSLRPSRPLALIEVAVAREPSASIGQLCTHIA 165
OY 254 EIG 256
DB 166 QLG 168
```

```

RESULT 10
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match          5.0%; Score 67; DB 15; Length 194;
Best Local Similarity 38.6%; Pred. No. 79;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

OY 202 AGNFIAIDKSGAGSRGSRG---IYALGPGG--KPSRIYIYTT 239
DB 75 SGRVTRPSGAGSRVTEGAEYRITAIAGPLAVHEPVAVAVVYT 118

RESULT 11
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvoegel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match          5.0%; Score 67; DB 15; Length 202;
```

Best Local Similarity 20.9%; Pred. No. 84;
Matches 39; Conservative 26; Mismatches 68; Indels 54; Gaps 7;

QY 50 LCCAVLSRIDAGEQGLGRIRHSQND--LWESPVTEKHL-----TDGNTVREL----- 97
DB 17 LLAAVLTFRQADVPVPRATRLPEAKDCHIAQFKSLPKELQAFKKAKAIERKLEKDM 76
QY 98 -CSA-----AITS--DNTAANLLTTTGGPKKEITAFIAN 129
DB 77 RCGSHLSRAMDKLQLOVQERPKALQAEVALTLKWNINDSALTTLIGQPLHLSHIS 136
QY 130 MGDHVT-----RIDRWEPELNEAIPNDEDTTPVAMATTLRLKLTGLTLL 176
DB 137 QLGCTQLQATAEPKPSRLSLRLHLQLQA--QSKETPGCLEDSVTSNLEQLLRLDKCV 195
QY 177 ASRQQL 183
DB 196 ASGDQCV 202

RESULT 12

US-09-764-868-757
Sequence 757, Application US/09764868
Patent No. US2002016871A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 757
LENGTH: 179
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-868-757

Query Match 4.8%; Score 65; DB 10; Length 179;
Best Local Similarity 27.3%; Pred. No. 1.1e+02;

Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

QY 120 PRELTAFLNMGDHYTRLD--RWEPELNEA-----IPNDEDTTPVAMATTLRLKLTG 171
DB 52 PRELTVV---QGEKLEVLDSKRWMLVKNAGRSGLTIPSNILEPQLPGTPTGOGSPSRV 108
QY 172 ELLTTLASR-QQLIDWMEADKVGAPLLRS 198
DB 109 PMLRLSSRPEEVTDWLQAEINFSTATVRT 136

RESULT 13

US-10-106-698-4858
Sequence 4858, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005P1
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 4858
LENGTH: 179
TYPE: PRT

ORGANISM: Homo sapiens
US-10-106-698-4858

Query Match 4.8%; Score 65; DB 15; Length 179;
Best Local Similarity 27.3%; Pred. No. 1.1e+02;

Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

QY 120 PRELTAFLNMGDHYTRLD--RWEPELNEA-----IPNDEDTTPVAMATTLRLKLTG 171
DB 52 PRELTVV---QGEKLEVLDSKRWMLVKNAGRSGLTIPSNILEPQLPGTPTGOGSPSRV 108
QY 172 ELLTTLASR-QQLIDWMEADKVGAPLLRS 198
DB 109 PMLRLSSRPEEVTDWLQAEINFSTATVRT 136

RESULT 14

US-10-189-346-12
Sequence 12, Application US/10189346
Publication No. US20030119162A1
GENERAL INFORMATION:
APPLICANT: Churchill, Mair
APPLICANT: von Bodman, Susanne
APPLICANT: Schweizer, Herbert
APPLICANT: Gould, Ty
APPLICANT: Hoang, Tung
APPLICANT: Murphy, Frank
APPLICANT: Watson, William
TITLE OF INVENTION: Structural Basis of Quorum Sensing Signal Generation and Metho
FILE REFERENCE: 2848-44
CURRENT APPLICATION NUMBER: US/10/189,346
CURRENT FILING DATE: 2002-07-02
PRIOR APPLICATION NUMBER: 60/303,449
PRIOR FILING DATE: 2001-07-04
PRIOR APPLICATION NUMBER: 60/366,575
PRIOR FILING DATE: 2002-03-21
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn version 3.1
SEQ ID NO 12
LENGTH: 202
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-189-346-12

Query Match 4.8%; Score 65; DB 15; Length 202;
Best Local Similarity 28.3%; Pred. No. 1.3e+02;

Matches 36; Conservative 18; Mismatches 41; Indels 32; Gaps 8;

QY 120 PRELTA-----FLNMGDHYTRLDRWE--PELNEAIPND--ERDTTPVAMATTLRK 167
DB 12 PYELAADLGRYRRRVFVQGL-----VALPSANAEFERDOPDRDRTTYVARNAGE 63
QY 168 LITGELLTLASRQQLIDWMEADKVG--PLIRSLPLPGWT-----ADRSAGG--ERGS 217
DB 64 MCGCARLPTTQPYLLSLFADIVAQDVLPKSA--AVWELSRPAATADENGCGNEMAV 121

QY 218 RGIITAA 224
DB 122 RPLMAAV 128

RESULT 15

US-10-156-761-8978
Sequence 8978, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIYA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_37S
Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARV...TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	39.0	105	2	JC2566 bla protein - Pseu
2	78	5.8	191	2	S67447 hypothetical prote
3	74	5.5	177	2	A83753 hypothetical prote
4	73.5	5.5	184	2	T21126 ADP-ribosylation f
5	72.5	5.4	113	2	T45195 hypothetical prote
6	72.5	5.4	200	2	G97064 spore coat protein
7	71.5	5.3	152	2	T36984 hypothetical prote
8	71.5	5.3	195	2	T36975 hypothetical prote
9	71	5.3	192	2	G83096 conserved hypotet
10	70.5	5.2	131	2	AD2281 conserved hypotet
11	70	5.2	108	2	E84217 hypothetical prote
12	69.5	5.2	145	2	F84251 hypothetical prote
13	69.5	5.2	167	2	D87360 conserved hypotet
14	69	5.1	180	2	C71859 conserved hypotet
15	68.5	5.1	172	2	AD3605 hypothetical prote
16	68.5	5.1	190	2	T09136 ADP-ribosylation f
17	68	5.1	170	2	AB0192 conserved hypotet
18	67.5	5.0	192	2	AB3587 conserved hypotet
19	67.5	5.0	198	2	D95285 conserved hypotet
20	66.5	4.9	128	2	F87353 conserved hypotet
21	65.5	4.9	442	2	F56978 collagen alpha 1(X
22	65.5	4.9	116	2	C82906 ribosome binding f
23	65.5	4.9	195	2	AE0623 conserved bacteriop
24	65.5	4.9	197	2	H90211 conserved hypotet
25	65	4.8	148	2	E75283 conserved hypotet
26	65	4.8	150	2	AC0284 probable heat shock
27	65	4.8	153	2	E81708 conserved hypotet
28	65	4.8	160	1	E69186 conserved hypotet
29	65	4.8	168	2	B75498 conserved hypotet

30	65	4.8	180	2	G70912 hypothetical prote
31	65	4.8	198	2	C69296 conserved hypotet
32	64.5	4.8	151	2	D81333 probable protein-t
33	64.5	4.8	177	2	D90227 hypothetical prote
34	64.5	4.8	188	1	G69365 conserved hypotet
35	64	4.8	113	2	D70580 hypothetical prote
36	64	4.8	142	2	A71000 hypothetical prote
37	64	4.8	150	2	T08585 calmodulin - soybe
38	64	4.8	162	2	AG0769 hypothetical prote
39	64	4.8	177	2	T48420 hypothetical prote
40	63.5	4.7	168	2	T20606 hypothetical prote
41	63.5	4.7	179	2	AB1994 hypothetical prote
42	63.5	4.7	193	2	S42528 GTP-binding protei
43	63.5	4.7	196	2	H69647 2-dehydro-3-deoxy-
44	63	4.7	122	2	AC3609 transcription regu
45	63	4.7	124	2	FC75373 conserved hypotet

ALIGNMENTS

RESULT 1

JC2566
bla protein - Pseudomonas aeruginosa (fragment)
C:Species: Pseudomonas aeruginosa
C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996
C/Accession: JC2566
R:West, S.E.H., Schweitzer, H.P., Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 14811281, 81-86, 1994
A>Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A>Note: due to a typographical error the volume number 148 appears as 128
A/Accession: JC2566
A:Molecule type: DNA
A:Residues: 1-105 <MES>
C:Genetics:
M:Gene: bla
C:Superfamily: beta-lactamase I

Query Match 39.0% Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 6.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	150	AMATTLRRLTGGELTTLASROOLDMEADRVAGPLRSALPAGMFTADKSGAERSRG	219
DB	2	AMATTLRRLTGGELTTLASROOLDMEADRVAGPLRSALPAGMFTADKSGAERSRG	61
QY	220	IIAALGPGRKSRIVYITTTGSGATMDERNRQIAEIGASLIKHW	263
DB	62	IIAALGPGRKSRIVYITTTGSGATMDERNRQIAEIGASLIKHW	105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: T38062; S67447
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z21766
A/Accession: T38062
A>Status: Preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-191 <MC2>
A:Cross-references: EMBL:Z69944; NID:91217974; PIDD:CAA93808.1; PID:91217978; GSPDB:G
C:Genetics:
A:Experimental source: strain 972n-; cosmid c1F12
A:Gene: SPAC1F12.04c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c
Query Match 5.8%; Score 78; DB 2; Length 191;

A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: G97064
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-200 <ORF>
 A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:G15024270; GSPDB:GN00168
 A:Experimental source: clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CACJ338

Query Match 5.4%; Score 72.5; DB 2; Length 200;
 Best Local Similarity 25.0%; Pred. No. 94;
 Matches 35; Conservative 27; Mismatches 55; Indels 23; Gaps 7;

OY 1 HPELVYKDAEDELGARGVETLDSKILSEFR-SEERFPMMSTFKVLLCGAVLSRI 59
 DB 10 HP---VAKIKPMNOL-AKVIITQYGGPDGELASIRLSQRFKSVTP-----QAIATLN 59
 OY 60 DAGOEQLGRRIHYSONDLVEYSPYTERKHLTDGMTVRELCSAALT--MSDNTANLLTTI 117
 DB 60 DIGTEELAH-----LEIVGSIVROLSRGLSVEELKSKGLDVFADHDSALYPSAA 110
 OY 118 GGPRELTAFLHNMGDHVTRL 137
 DB 111 GNP-FTAAYIOSKGDPTTDL 129

RESULT 7

T36984

hypothetical protein SCJ11.13 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000

C:Accession: T36984

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1999

A:Reference number: Z21618

A:Accession: T36984

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-152 <OL1>

A:Cross-references: EMBL:AL109949; PIDN:CA852898.1; GSPDB:GN00070; SCOEDB:SCJ11.13

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCJ11.13

C:Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
 Best Local Similarity 23.4%; Pred. No. 79;
 Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

OY 106 DNTANLLTTTGGPRELTAFLHNMGDHVTRLDRWPELNEAIPNDRDTPVAMATTL 165
 DB 25 DDAAMATTLAQVTPPTPADLTFYESIGD-VT---MEDVNGVFNP----- 65
 OY 166 RKLITGELLTLASROQLIDMMEDAKVAGPLLSALPAGFTADKSGAGSGRGITIALG 225
 DB 66 ----AGDLRLQEGGVWDEGTDEKSRGLVT-----GSMNG-GLIYVAG 104
 OY 226 PGKPSRIYVITTSQATMDERN-----RQIAEI 255
 DB 105 PDG-----AVYRTRT-ASLDEALDKVADGIRQLEL 135

RESULT 8

T36975

hypothetical protein SCJ11.04 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36975

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1999

A:Reference number: Z21618

A:Accession: T36975

A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-195 <OL1>
 A:Cross-references: EMBL:AL109949; PIDN:CA852889.1; GSPDB:GN00070; SCOEDB:SCJ11.04
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCJ11.04

Query Match 5.3%; Score 71.5; DB 2; Length 195;
 Best Local Similarity 17.6%; Pred. No. 11e+02;
 Matches 34; Conservative 35; Mismatches 63; Indels 61; Gaps 7;

OY 56 LSRIDAGOEQLGRRIHYSONDLVEYSPYTERKHLTDGMTVRELCSAALTMSDNTANLLTTI 115
 DB 46 VALLENGEIEFLDRV-----IAREELVLAASAETGVTAEAGEGELALVP 93
 OY 116 TIGPRELTAFLHNMGDHVTRLDRWPELNEAIPNDRDTPVAMATTLRKLITGEL 173
 DB 94 APAASAEPGAI-----VPHQEGLSVSLSPNNQR-----I 124
 OY 174 LTLASROQLIDMMEDAKVAGPLLSALP-----GFTADKSGAGSGRSR 218
 DB 125 LNTLDPRPGLEPYRADIDAAALGTEAAAKVGEVRPKRLAERGMILQEAAGAFSAGR 184
 OY 219 GIIALGPDGKPS 231
 DB 185 -LVAS--PGDPS 194

RESULT 9

G83096

conserved hypothetical protein PA4399 [imported] - Pseudomonas aeruginosa (strain PAO)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001

C:Accession: G83096

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa

thogen; PMID:10964043

A:Reference number: AB2950; MUID:20437337; PMID:10964043

A:Accession: G83096

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-192 <STO>

A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA07787.1; GSPDB:GN

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA4399

C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
 Best Local Similarity 26.0%; Pred. No. 1.2e+02;
 Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

OY 53 GAVLSRI-----DAGOEQL--GRRIHYSONDLVEYSPYTERKHLTDGMTVRELCSA----- 100
 DB 2 GNRUSKITRTGDSGETGLAGRRVPSHPRIEAIQAVDELNSQLGLLLELLEARAH 61
 OY 101 -----AITMSDNTANLLTTTGGPRELTAFLHNMGDHVTRL-----DWREPL-----N 145
 DB 62 GLEIYVALAPVGR-----LFDLGGELAMPEYRALDETEVARLESCIDRWNNDELGPLKN 116
 OY 146 EAIRPDERDTPVAMATTLRKLITGELLTLASROQLIDMMEDAKVAGPLLSALPAGWF 205
 DB 117 FILRGSR-----PAQAHVCRSLAR-----SAERRCQALDDEETLEGGVGLVLRSLSLP 168
 OY 206 IADKSGAGSGRGII--AALGPD 227
 DB 169 VAAAIARAGGVATILMEAAAKPD 192

RESULT 10

AD2281
 hypothetical protein alr3803 [Imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD2281
 R:Kanehisa, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matsubae, A.; Iritani, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2281
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-131 <STO>
 A:Cross-references: GB:BA000019; PIDN:BAF75502.1; PID:q17132937; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr3803

Query Match 5.2%; Score 70.5; DB 2; Length 131;
 Best Local Similarity 26.0%; Pred. No. 78;
 Matches 34; Conservative 19; Mismatches 51; Indels 27; Gaps 5;
 QY 77 LVESPYTEKHLTDOMTVRELCSAA-ITMSDNTAANLLTTIGPKETAFLLHNM--GDH 133
 DB 1 MVMSSVTLQNDVLTFLFSSQNCOLYDIOTEPFLDLPYYSKFNRRTALVGNF 60
 QY 134 VTRLD--RWEPELNEAIPNDERDTPVMAATTLRKLLT-----GE 172
 DB 61 VISEFWTMMODEANFVFNPE---LNLADISTIKLTSRKERKRCGSLAQMIDNGH 117
 QY 173 LITLASROOLI 183
 DB 118 LITLQRLQAI 128

RESULT 11
 E84217
 hypothetical protein Vng0594h [Imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84217
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jaldic
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483; PMID:11016950
 A:Accession: E84217
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA619105.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG0594H

Query Match 5.2%; Score 70; DB 2; Length 108;
 Best Local Similarity 27.5%; Pred. No. 66;
 Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;
 QY 44 MSTFKVLLCGAVLSRIDAGOEGLRRIHYSQNDLVEYSPTEKHLTDGMTVELCSAAIT 103
 DB 1 MDTTVELDQALLARLEARD-----RVFVRDPALEVTDTLRFRRHGDGV-----GSIT 50
 QY 104 MSDNTAANLLTTIGPKEL-----TAFLLHMGCHVTFLDR 139
 DB 51 NDDGTDTMARLTPGDSDFIAVEPTSEVAALVDAATRTDR 92

RESULT 12
 F84251

hypothetical protein Vng0953c [Imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84251
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483; PMID:11016950
 A:Accession: F84251
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <STO>
 A:Cross-references: GB:AE004437; NID:g10580510; PIDN:AA619378.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG0953C

Query Match 5.2%; Score 69.5; DB 2; Length 145;
 Best Local Similarity 20.3%; Pred. No. 11e+02;
 Matches 30; Conservative 29; Mismatches 66; Indels 23; Gaps 3;
 QY 41 FPMKSTFKVLLCGAVLS-----RIDAGEGLRRIHYSQNDLVEYSPTEKH 87
 DB 3 FSVSGSAALIFIAFVSVGLYSAPNGFERVODANNANDRYLTKNTVEAANTTYS 62
 QY 88 LITDOMTVRELCSAAITMSDNTAANLLTTIGPKETAFLLHMGCHVTFLDRWEPELNEA 147
 DB 63 VNDTVYMAATNNKSTLISVQTDVY---DGEVYDSAVYSSVDCNSQTDLM----- 112
 QY 148 IPNDERDTPVMAATTLRKLLTGLLELT 175
 DB 113 LPGETSVTVPTGSAPTPRVKTGTGT 140

RESULT 13
 D87360
 conserved hypothetical protein CC0895 [Imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: D87360
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Debby, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: AB7249; MUID:21173698; PMID:11259647
 A:Accession: D87360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-167 <STO>
 A:Cross-references: GB:AE005673; NID:g13422160; PIDN:AAK22880.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0895

Query Match 5.2%; Score 69.5; DB 2; Length 167;
 Best Local Similarity 23.4%; Pred. No. 1.3e+02;
 Matches 36; Conservative 19; Mismatches 56; Indels 31; Gaps 6;
 QY 2 PETLVKVKDA-----EDQLGARVGYIELDN---SGKILSEFRSEERPPMSTFK 48
 DB 24 PHTLLILDPAFAGYKRFQOMQDRICGVARNVLAALKTLVSHGVETRRYSER-PPRIEYW 82
 QY 49 VILCGAVLSRIDAGOEGLRRIHYSQNDLVEYSPTEKHLTDGMTVR-----ELCSAAIT 103
 DB 83 LTERKIALSPVLLTVAEWGDRHYVGR---DKSPVLEFRHKYCGCAFHVPLVACAGGVVD 138
 QY 104 MSDNTAANLLTTIGPKELTA 125
 DB 139 RRDIERA-----GPDLLTA 152

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_37S
Perfect score: 1345
Sequence: 1 HPEITLVKVADEADQAGARVG.....TMDERNQIAETGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	5.8	191	YDA4_SCHPO	Q10346 schizosacch
2	69.5	5.2	196	RAC2_LOTJA	Q40220 lotus japon
3	69.5	5.1	201	RACG_DICDI	Q99850 dictyostell
4	68	5.1	122	HC3L_THIFE	P80509 thioabactill
5	65.5	4.9	116	RBRA_UREPA	Q99490 ureaplasma
6	65.5	4.9	206	KTHY_METAC	Q87959 methanosarc
7	65	4.8	198	OGGI_ARCFU	Q29876 archaeoglob
8	63.5	4.7	193	SAR2_LYCES	P52884 lycopersico
9	63.5	4.7	196	ALRH_BACSU	P50846 b.khg/kdpg
10	63.5	4.7	200	RR4_PELNE	Q95459 pellicla nees
11	63	4.7	176	HSIV_THEMA	Q94721 thermotoga
12	62.5	4.6	144	MARR_ECOLI	P27245 escherichia
13	62.5	4.6	146	HBC_RABIT	P02099 oryctolagus
14	62.5	4.6	160	TATB_HELPY	Q25700 helicobacte
15	62.5	4.6	172	YDEJ_ECOLI	P31131 escherichia
16	62.5	4.6	182	PYRE_STRCO	Q9487 streptomyc
17	62.5	4.6	184	HRPL_PSESY	P37929 pseudomonas
18	62.5	4.6	195	TRPE_THEYO	Q97976 thermoplas
19	62	4.6	149	NCK_TREPA	Q85974 treponema p
20	62	4.6	173	Y265_BORBU	Q51280 borrelia bu
21	62	4.6	174	IHBB_RAT	P17491 rattus norv
22	62	4.6	182	RR4_BEICH	O19990 belamanda
23	62	4.6	182	Y861_BIFLO	O86520 bifidobacte
24	61.5	4.6	184	MURL_SCHPO	Q9485 schizosacch
25	61.5	4.6	185	YKQC_ECOLI	P38522 escherichia
26	61.5	4.6	197	HANI_XENLA	Q73615 xenopus lae
27	61	4.5	177	VNSC_RINDK	P35948 rinderpest
28	61	4.5	177	VNSC_RINDK	Q03339 rinderpest
29	61	4.5	184	ARL2_DROME	Q06849 drosophila
30	61	4.5	194	RR4_IIRIN	O20234 iritis ensata
31	60.5	4.5	126	YGM1_YEAST	P53130 saccharomyc
32	60.5	4.5	142	Y863_METTM	Q50770 methanobact
33	60.5	4.5	150	PDOV_SALTY	Q94xdm6 salmoneilla

34	60.5	4.5	152	YU33_YERPE	O86c68 yeastinia pe
35	60.5	4.5	160	TATB_HELPY	Q92m58 helicobacte
36	60.5	4.5	178	HSIV_RALSO	Q8y3d7 raistonia s
37	60	4.5	121	SECR_HUMAN	P09683 homo sapien
38	60	4.5	161	PIN_BP74	P07068 bacterioph
39	60	4.5	178	PYRE_ARCFU	O28533 archaeoglob
40	60	4.5	178	UCRI_CANCP	P70758 anabaena sp
41	60	4.5	200	TATB_CANCP	Q96c1 caulobacter
42	60	4.5	203	HOP2_YEAST	P53187 saccharomyc
43	59.5	4.4	103	RS10_NBIGO	P48851 nelsieria g
44	59.5	4.4	156	BFR_AZOVI	P22759 azotobacter
45	59.5	4.4	177	ATPD_HAEIN	P43717 haemophilus

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD: PRT: 191 AA.
ID YDA4_SCHPO
AC 010346:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Clf12.04c in chromosome I.
GN SPAC1F12.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972:
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Iyne M., Iyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Nibbelk D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weljens I., Vanslreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Flitz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck C., Holzer E., Moestl D., Hilbert H.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardo L., Jimenez J., Sanchez A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez A., Thode G.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Paulsen I., Potashkin J.,
RT *The genome sequence of Schizosaccharomyces pombe.*;
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: 269944; CA93808.1; -
DR PIR: T38062; S67447
DR GenedB_Spombe; SPAC1F12.04C; -

OY 44 MSTFKVLLGAVLSRIDAGEQLGRTHYSQNDLV-EXSP-VTEKHLTDGMYRELCSAA 101
 DB 1 MKSITKCVVG-----EGGIGKTSKMLSTYSNSISNISOPTVDNST----- 42
 OY 102 ITNSDNTAANLLTTTGPKELTAFA-----LHNMGDHYRLDRWEPELN 145
 DB 43 LLMHNKFPYLSLMDPAGGEFSEKRLSLSPQDVPILCSLNPSPFSNILLSDWOELN 102
 OY 146 EALPNDERTMVPVAAATLTKLLTGELL--TLASRQOL 182
 DB 103 ENCPN-----TPVLVGTOMDLKSNVILDRLECKKQL 135

RESULT 4

HC3L_THIFE STANDARD: PRT: 122 AA.
 ID HC3L_THIFE
 AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase) (fragments).
 GN HOXG.
 OS Thiobacillus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 OX NCBI_TaxID=920;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=ATCC 19659;
 RX MEDLINE=96241862; PubMed=8661919;
 RA Fischer J., Quentmeier A., Kostka S., Kraft R., Friedrich C.G.:
 RT Purification and characterization of the hydrogenase from
 RL Thiobacillus ferrooxidans.
 CC Arch. Microbiol. 165:289-296(1996).
 CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 - 2 H(+) + 2
 CC ferredoxin.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 KM Oxidoreductase, iron.
 FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 76 77
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EBAFAF65348DA0B CRC64;

Query Match 5.1%; Score 68; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 73;
 Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

OY 19 VETIYIELDNSGKILSFSEEPFPMKSTFKVLLCGAVLSRIDAGEQLGRTHYSQNDLV 78
 DB 12 VERVEGDLDLSKSLIEFFRN-----AIIARFGCG-----LG 41
 OY 79 EYSPYTERKHLTDGMYRELCSAATMDSNTAANLLTTTGPKELN 124
 DB 42 KAAPFTGTYVEIGV-----ISGDKDPQAGLVV-----PREST 75

RESULT 5

RBFA_UREPA STANDARD: PRT: 116 AA.
 ID RBFA_UREPA
 AC O9POH0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ribosome-binding factor A.
 GN RBFA OR U1321.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
 OX NCBI_TaxID=134821;
 RN [1]

SEQUENCE FROM N.A.

RP STRAIN=Serovar 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.D., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Cassell G.H.;

RT "The complete sequence of the mucosal pathogen Ureaplasma
 urealyticum."
 RL Nature 407:757-762(2000).

CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
 with 30S subunits that are part of 70S ribosomes or polysomes).
 CC Essential for efficient processing of 16S rRNA. May interact with
 CC the 5' terminal helix region of 16S rRNA (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AE002129; AAF30730.1; -
 CC HAMAP: MF_00003; -; 1.
 CC DR InterPro: IPR000238; Rib_bind_facta.
 CC DR Pfam: PF02033; RBFA; 1.
 CC DR ProDom: PD007327; Rib_bind_facta; 1.
 CC DR TrGPro: TRG00082; TbfA; 1.
 CC DR PROSITE: PS01319; RBFA; FALSE_NEG.

DR RNA processing: Complete proteome.
 SK SEQUENCE 116 AA; 13247 MW; A79DC71F0547514 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No. 11e+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

OY 144 LNEAIPNDRDTMPVAAATLTKLLTGEL-----LTLASRQOLDMKE-ADRYAGPL 196
 DB 18 INNALANEINDKIARATYAVR--LSNLSVAKIFLDHAKHESMLKLVLENVAKVSG-LL 74

OY 197 RSALPAGW-----FIADKS 210
 DB 75 RSKLAEMWTSYKVPRLRFVIDET 97

RESULT 6

KTHY_METAC STANDARD: PRT: 206 AA.
 ID KTHY_METAC
 AC Q8THS9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN TMK OR M44433.
 OS Methanosaetia acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,

RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kmetzner J.A., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umagay L.A., White O., White A.J.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Zander E.,
 RA Metcalf W.W., Birren B.;
 RA "The genome of *Methanocaldococcus jannaschii* reveals extensive metabolic
 RT and physiological diversity";
 RL Genome Res. 12:532-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate - ADP + thymidine
 CC 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE011164; AAM0774.1; -
 DR HAMAP; MF_00165; -; 1.
 DR InterPro; IPR000062; Thymidylate_kin.
 DR Pfam; PF02223; Thymidylate_kin; 1.
 DR TRIGRAMS; TIGR00041; DIMP_kinase; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
 KM Transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
 KM Complete proteome.
 FT NP_BIND 10 17 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 23303 MW; A19C25BC85423B3 CRC64;
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.2e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 QY 56 LSRID-AGQELGRIRHYSQNDLYEYSPYTEKHLTDGTYRELCSAATMSDNTAANLL 114
 DB 8 LEGIDGSKSVAKKIQ-ENSELRFVEVFTREPTRG-----TLGNMVENAIQ 55
 QY 115 TTIGCPKELTAFLLNMGDVTRLDWEPELNEAIPNDERBOTTVPVMAATITRLTGELL 174
 DB 56 SDTQDLAELEFTADHAEHLAKLVKPALEDGRTVSDYSRVAOGITLKNRLDPL- 114
 QY 175 TLASRQQLIDWMEADKVAQPLRSALPAGW-----FIAD-----KSG-AGE----- 214
 DB 115 -----EWR-----DLRGMVITDITFLPIEPIEIAVKRCGRKEQTKF 154
 QY 215 -----RGSRGIIALGPGRPSRIIVYTTGSAQATMDE 247
 DB 155 EKIEFLRGVRELEFLGLAE-EPEPVIVDASGSPEDVEK 192
 RESULT 7
 OGG1_ARCFU STANDARD; PRT; 198 AA.
 ID OGG1_ARCFU 029876;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
 DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
 DE (EC 4.2.99.18) (AP lyase)].
 GN OGC OR AF0371.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID-2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VQ-16 / DSM 4304 / ATCC 49558;

RX MEDLINE-98049343; PubMed-9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Kirschum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kervagay A.R., Graham D.E., Kyriades N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodex A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*";
 RL Nature 390:364-370(1997).
 CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
 CC guanine (7,8-dihydro-8-oxoguanine - 7-oxoG) from DNA. Also nicks
 CC DNA at apurinic/apyrimidinic sites (AP sites) (by similarity).
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
 CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
 CC leaving a 3'-terminal unsaturated sugar and a product with a
 CC terminal 5'-phosphate.
 CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001079; AAB90876.1; -
 DR PIR; C69296; C69296.
 DR TIGR; AF0371; -
 DR InterPro; IPR003265; Endo_3c.
 DR HAMAP; MF_00241; -; 1.
 DR TIGR; AF0371; -
 DR Pfam; PF00730; Hnh-GPD; 1.
 DR SMART; SM00478; ENDO3C; 1.
 KM Hydrolyase; Nuclease; Endonuclease; Lyase; DNA repair; glycosidase;
 KM Multifunctional enzyme; Complete proteome.
 FT ACN_SITE 122 122 BY SIMILARITY.
 SQ SEQUENCE 198 AA; 22639 MW; 3A5C033AA1233FFB CRC64;
 Query Match 4.8%; Score 65; DB 1; Length 198;
 Best Local Similarity 24.6%; Pred. No. 2.3e+02;
 Matches 51; Conservative 25; Mismatches 85; Indels 46; Gaps 14;
 QY 14 QLGARVGIETLDL---SGRLIESFRSEDERPPMASTFKVLLCGAVLSRIDAGQ-----EQ 65
 DB 15 QLGKR-GEVERDFRPFIDFSVKATIRTELAFICISTANSSATAGLKFQRL-IGQGVKKA 72
 QY 66 L---GRRIHYSQNDLYEYSPYTEKHLTDGMYRELCSAATMSDNTAANLLTTIG-GPK 121
 DB 73 LTLAGVRFH---NKAET-----IREAPSPFLVERKALEABESSKREIILTKIGLGMK 122
 QY 122 ELTAFLLNMG-DHYTRLD---RWEPELNEAIPNDERBOTTVPVMAATITRLT-----T 170
 DB 123 EASHFLRVGREDVAIIDRHILRWLERQGYEVP-----TWTAKVLEVERILMEISEER 177
 QY 171 GELLTFLASRQQLIDWMEADKVAQPLR 197
 DB 178 GESL---AEMDLRIWAE---MTGVLK 198
 RESULT 8
 SAR2_LYCES STANDARD; PRT; 193 AA.
 ID SAR2_LYCES P52884;
 AC 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE GTP-binding protein SAR2.

KHG/KDPG aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase (EC 4.1.3.10) (2-keto-4-hydroxyglutarate aldolase) (KHG-aldolase); 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate aldolase)] (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-aldolase)].

KDGA.

Bacillus subtilis.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NCBI_TaxID=1423;
(1)

SEQUENCE FROM N.A.
RX STRAIN=168 / Marburg;
MEDLINE=96349105; PubMed=8760912;
Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
Serror P.;
Sequence analysis of the Bacillus subtilis chromosome region between
the serA and kds loci cloned in a yeast artificial chromosome.";
Microbiology 142:2005-2016(1996).
(2)

SEQUENCE FROM N.A.
RC STRAIN=168;
MEDLINE=98044033; PubMed=9384377;
Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borissoff L., Bourcier L., Brans A., Braun M., Brignelli S.C., Bron S.,
Broillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Enlita K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kashnira Y., Klaerr-Bianchand M., Klein C.,
Kobayashi Y., Koester P., Kondigstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
Parro V., Pohlt T.M., Portecelle D., Porwollik S., Prescott A.M.,
Prescan E., Pujeat P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takenouchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognotti A.,
Toledo V., Uchiyama S., Vandendool M., Vanlier F., Vassarotti A.,
Vialti A., Wambut R., Wedler E., Wedler H., Weissmeyer T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumschein E., Yoshitawa H., Danchin A.;
The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis ";

Nature 390:249-256(1997).

-1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate = pyruvate +
glyoxylate.

-1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate =
pyruvate + D-glyceraldehyde 3-phosphate.

-1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.

-1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
OF GLYOXYLATE.

-1- SUBUNIT: Homotrimer (By similarity).

-1- SUBCELLULAR LOCATION: CYTOPLASMIC.

-1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
institutes requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@lsb.sib.ch).

```

DR EMBL: LA7838; AAB38480.1; -
DR EMBL: Z99115; CAB14127.1; -
DR PIR: H69647; H69647.
DR Subtilisin: BG1396; kda.
DR InterPro: IPR000887; Aldise_KDPG_KMG.
DR Pfam: PF01081; Aldolase; 1.
DR TrGFams: TIGR01182; ada; 1.
DR PROSITE: PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
DR PROSITE: PS00160; ALDOLASE_KDPG_KMG_2; 1.
DR Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
FT ACT_SITE 43 43 BY SIMILARITY.
FT ACT_SITE 47 47 BY SIMILARITY.
FT BINDING 130 130 SCHIFF-BASE WITH KMG OR PYRUVATE (BY SIMILARITY).
SQ SEQUENCE 196 AA; 20865 MW; EB0B597FD1102CC4 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 196;
Best Local Similarity 22.1%; Pred. No. 3.1e+02;
Matches 27; Conservative 17; Mismatches 59; Indels 19; Gaps 4;

QY 31 ILSEFRSEERFPMKSTFKVLLCGAVLSRIDAGQ-EDGRIHNSON---DLVEYSPYTEK 86
DB 54 ITSEFRREDI-----LIGAGTVLSAQAGEMAKAGQIVYSPGSADLAELSPVK 106
QY 87 HLTGKMTVREICSAIITMSDNTAANT-----LTTIGPKELTAFLNMGDHYTRLD 138
DB 107 HYICGVLTPEKIMALVFEGFTTLKLPSPGVGFIPEMKNLGAPPROVTFEPTGSIHSEVP 166
QY 139 RW 140
DB 167 DW 168

RESULT 10
RR4_PELNE STANDARD; PRT: 200 AA.
AC Q9PSD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Peilla neesiana (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellineae;
OC Pellineae; Pellia.
OX NCBI_TaxID=70144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gametophyte;
RA Capelus I., Bloecher R.;
RT "A molecular approach to bryophyte systematics.";
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: One of the primary rRNA binding proteins. It binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -1- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send pm email to license@isb-sib.ch).

```

```

CC EMBL: AJ250456; CAC14050.1; -
DR HSSP: P81288; 1C05.
DR HAMAP: MF_01306; 1.
DR InterPro: IPR001912; Ribosomal_S4.
DR InterPro: IPR002942; S4.
DR InterPro: IPR005709; S4_bact_org.
DR Pfam: PF00163; Ribosomal_S4; 1.
DR Pfam: PF01479; S4; 1.
DR SMART: SM00363; S4; 1.
DR TrGFams: TIGR01017; rpsd_bact; 1.
DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
DR PROSITE: PS00889; S4; 1.
DR Ribosomal protein; RNA-binding; Chloroplast.
FT DOMAIN 90 150 S4 RNA-BINDING.
FT SEQUENCE 200 AA; 23178 MW; 00D0098C6D9933F5 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 200;
Best Local Similarity 22.8%; Pred. No. 3.1e-02;
Matches 39; Conservative 31; Mismatches 78; Indels 23; Gaps 7;

QY 14 QGARYGYIE--LDLNSGKILSESEERFPMKSTFKVLLCGAVLSRIDAGQGLRRYH 71
DB 14 RIGALPGLSKILELSEGIIGSTPKR-----VSQYRIKL-----EKGKLPFH 58
QY 72 Y--SQNDLVEYSPYTEKHLTDGTVRELCSAIITMSDNTAANT-LTTIGPKELTAFLH 128
DB 59 YGILERQLLKLYVIARK--AKSGTCQIILSOTLEMRDNIIFRIGMSPTIPGARGLVNRH 116
QY 129 NM-GDHYTRLDREPELNEAIPDERDTTPVAMATTLKRLTGLLETLAS 178
DB 117 ILINDTVDPINCEPKDVIYNNKESVIITKNMDSKRPVPHNLTDS 167

RESULT 11
HSLV_THEMA STANDARD; PRT: 176 AA.
AC Q3WIZ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease hslV (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109;
RX MEDLINE=99287316; Pubmed=10360571;
RA Hail D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uitterlind T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send pm email to license@isb-sib.ch).

```

DR EMBL; AE001728; AAD35606.1; -.
 DR PTR; G72365; G72365.
 DR HSSP; P31059; IE94.
 DR MEROPS; T01.006; -.
 DR TIGR; TM0521; -.
 DR HAMAP; MF_00248; -. 1.
 DR InterPro; IPR001355; Protsme_protease.
 DR Pfam; PF00227; Proteasome; 1.
 KW Hydrolase; Protease; Complete proteome.
 FT ACT_SITE 6 6
 SQ SEQUENCE 176 AA; 18933 MW; EC635602A0AD02 CXC6;

Db 105 AAICBCHOLVGDHQLHQLTKLTADEVATLEYLKKVLP 144

RESULT 13
HBC_RABIT
ID HBC_RABIT STANDARD: PRT: 146 AA.
AC P02039;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin gamma chain (beta-3).
GN HBG.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82053017; PubMed=6271761;
RA Hardison R.C.;
RT "The nucleotide sequence of rabbit embryonic globin gene beta 3.";
RL J. Biol. Chem. 256:11780-11786(1981).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89178632; PubMed=2486295;
RA Margot J.B., Demers G.W., Hardison R.C.;
RT "Complete nucleotide sequence of the rabbit beta-like globin gene cluster. Analysis of intergenic sequences and comparison with the human beta-like globin gene cluster.";
RL J. Mol. Biol. 205:15-40(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN.
CC -1- SUBUNIT: Heterodimer of two alpha chains and two gamma chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: M18818; AAA02984.1; -;
CC DR EMBL: V00883; CAA2452.1; -;
CC DR PIR: A02417; HBRB3.
CC DR HSSP: P02100; 1A9W.
CC DR InterPro: IPR003337; Beta_haem.
CC DR InterPro: IPR003971; Globin.
CC DR Pfam: PF00042; Globin; 1.
CC DR PRINTS: PRO0814; BETAHAEM.
CC DR PROSITE: PS01033; GLOBIN; 1.
CC KW Heme; Oxygen transport; Transport; Erythrocyte; Embryo.
CC FT INIT_MET 0
CC SEQUENCE 146 AA; 16093 MW; EB8D6C1C24DD2D82 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 146;
Best Local Similarity 30.8%; Pred. NO.2.5e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

OY 35 FRSEERPPMSTFVKVLCGAVLSRIDAGQOLGRIRHSQNDLYEYSVTEKHLTDGNTV 94
DB 3 FTAEBKAATITSTWK-----LVVDADAGAEALGRLE-----LVVY-FWTFRFDSPGNTL 48

OY 95 RELCSAATITMSDNTAANLLTTTGGPKELTAFLHMGDHYTRLD 138
DB 49 SS-SSAIMGNPKNRAH-----GKRVLTAF-----GDAVKNVD 79

RESULT 14
TATB_HELPY

ID TATB_HELPY STANDARD: PRT: 160 AA.
AC 025700;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatB homolog.
GN TATB OR HPI060.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OC NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kershawage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter pylori.";
RL Nature 388:539-547(1997).
CC -1- FUNCTION: Required for correct localization of precursor proteins bearing signal peptides with the twin arginine conserved motif S/T-R-R-X-Y-L-K. This sec-independent pathway is termed TAT for twin-arginine translocation system. This system mainly transports proteins with bound cofactors that require folding prior to export (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC -1- SIMILARITY: Belongs to the tatB family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000613; MAD08110.1; -;
CC DR PIR: D64652; D64652.
CC DR TIGR: HPI060; -;
CC DR HAMAP: MF_00237; -; 1.
CC DR InterPro: IPR003369; Mta_Hcf106.
CC DR InterPro: IPR003998; TatB.
CC DR Pfam: PF02416; Mta_Hcf106; 1.
CC DR PRINTS: PRO1506; TATBPROTEIN.
CC DR TIGRFAMs: TIGR01410; tatB; 1.
CC KW Transport; Protein transport; Translocation; Transmembrane;
CC KW Inner membrane; Complete proteome.
CC FT TRANSMEM 1
CC SEQUENCE 160 AA; 18332 MW; 3BD4423AF7ABBD7 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 160;
Best Local Similarity 20.5%; Pred. NO.2.8e+02;
Matches 27; Conservative 30; Mismatches 64; Indels 11; Gaps 5;

OY 35 FRSEERPPMSTFVKVLCGAVLSRIDAGQOLGRIRHSQ--NDLYEYSVTEKHLTD-- 90
DB 19 FLGPEKPPQAVVDVKKFRFVAVKTKTLNDAKDTLDLDEIEIKETLEYQKLFENKVESLK 78

OY 91 GMTARELSAATITMSDNTAANL--LTTTGGPKELTAFLHMGDHYTRLDREPELEAI 148
DB 79 GVKIEELEDKAVT-AENKISIQDLMQYOKSLFTNTPIRHNLDEYSN-----EALNKREY 133

OY 149 PNDEEDTTPYA 160
DB 134 SSDESPEVQLA 145

```
RESULT 15
YDEJ_ECOLI STANDARD; PRT; 172 AA.
ID YDEJ_ECOLI
AC P31131;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ydeJ.
GN YDEJ OR B1537.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186717; PubMed=8383113;
RA Cohen S.P., Haechler H., Levy S.B.;
RT "Genetic and functional analysis of the multiple antibiotic
RT resistance (mar) locus in Escherichia coli."
RL J. Bacteriol. 175:1484-1492(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RC MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horluchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
CC -! SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E. COLI YGAD.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M96335; -; NOT_ANNOTATED_CDS.
DR EMBL; AE000251; AAC74610.1; -
DR EMBL; D90796; BAA15227.1; -
DR EMBL; D90797; BAA15339.1; -
DR PIR; D64908; D64908.
DR Ecogene; EG11645; ydeJ.
DR HAMAP; MF_00226; atypical; 1.
DR InterPro; IPR003396; CINA.
DR Pfam; PF02464; CINA; 1.
DR TIGRfams; TIGR00199; cina_cterm; 1.
RW Hypothetical protein: Complete proteome.
SQ SEQUENCE 172 AA; 18321 MM; 700A7A95C967C868 CRC64;
Query Match 4.66; Score 62.5; DB 1; Length 172;
Best Local Similarity 27.18; Pred. No. 3.1e+02;
Matches 23; Conservative 9; Mismatches 26; Indels 27; Gaps 3;
```

```
OY 73 SQNDLYEYSPYTEKHLTDGTVRELCSAITSNDNTANLLLTIGP----- 120
Db 79 SQOSLERYSAVSEK-----VAAEMATGAIERAD-ADVSIATITGCGPEGGEDTPAGTV 131
OY 121 -----KELTAFLEHMGDHYTRL 137
Db 132 WFAWHIKGQNTYAVMHFAGDCETVL 156
```

Search completed: September 10, 2003, 12:20:49
Job time : 5.57143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_37S
Perfect score: 1345
Sequence: 1 HPERLVKVKNAEDQLGARVC.....TMDERNQIAIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_23:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.minc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.potent:*
13: sp.virus:*
14: sp.vertebrate:*
15: sp.unclassified:*
16: sp.virus:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	39.0	105	2	Q52026
2	524	39.0	105	2	Q52330
3	433	36.7	145	2	Q8RTP8
4	488	36.3	102	2	Q52639
5	458	34.1	138	2	Q990F3
6	458	34.1	138	2	Q9AMA1
7	458	34.1	139	2	Q9AMA2
8	456	33.9	138	2	Q9AMA9
9	456	33.9	138	2	Q9AMA9
10	453	33.7	139	2	Q9AMA3
11	338	25.1	67	2	Q53553
12	320	23.8	95	2	Q8GDES
13	320	23.8	95	2	Q8GDES
14	316	23.5	62	2	Q9UN58
15	299	22.2	128	2	Q8VQ00
16	245	18.2	180	2	Q8KVT2

17	242	18.0	52	2	Q9R4I2	Q9R4I2 shigella fl
18	241	17.9	48	2	Q9RLH0	Q9RLH0 proteus mir
19	197	14.6	38	2	P97145	P97145 escherichia
20	173.5	12.9	134	2	Q8VTL3	Q8VTL3 staphylococ
21	128.5	9.6	109	2	Q53698	Q53698 staphylococ
22	105	7.8	20	2	P97146	P97146 escherichia
23	87	6.5	100	2	Q93505	Q93505 staphylococ
24	82	6.1	68	2	Q9XBJ2	Q9XBJ2 bacillus ce
25	79	5.9	188	2	Q9ACM8	Q9ACM8 streptococ
26	78	5.8	153	5	Q8SSZ0	Q8SSZ0 clona intes
27	77	5.7	181	15	Q8D108	Q8D108 yersinia pe
28	76.5	5.7	202	16	Q98520	Q98520 rhizobium i
29	74	5.5	177	16	Q9XEM7	Q9XEM7 bacillus ha
30	73.5	5.5	184	5	Q45379	Q45379 caenorhabdi
31	73	5.4	175	17	Q8RTM8	Q8RTM8 pyrobaculum
32	72.5	5.4	143	2	Q49970	Q49970 methanocarc
33	72.5	5.4	115	2	Q05984	Q05984 mycobacteri
34	72.5	5.4	200	16	Q97JEB	Q97JEB staphylococ
35	71.5	5.3	152	16	Q9RI91	Q9RI91 streptomyc
36	71.5	5.3	195	16	Q9RI91	Q9RI91 streptomyc
37	71.5	5.3	131	2	Q9X9H0	Q9X9H0 yersinia en
38	71	5.3	169	17	Q972W2	Q972W2 sulfoblob
39	71	5.3	192	16	Q9HW07	Q9HW07 pseudomonas
40	71	5.3	203	4	Q96BZ3	Q96BZ3 homo sapien
41	71	5.3	131	16	Q8YQ18	Q8YQ18 anabaena sp
42	70.5	5.2	108	17	Q9HRO3	Q9HRO3 halobacteri
43	70	5.2	150	10	Q94IG4	Q94IG4 nicotiana t
44	70	5.2	108	10	Q94IG4	Q94IG4 nicotiana t
45	70	5.2	205	6	Q9N275	Q9N275 ovis aries

ALIGNMENTS

RESULT 1

Q52026 PRELIMINARY; PRT; 105 AA.
ID 052026
AC 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DI 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN bla.
OS Pseudomonas aeruginosa.
OG Plasmid pR01614.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBL_TaxID=287;
RN [1]
RP
RX MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweizer H.P., Dail C., Runyen-Janecky L.J.;
RT "Construction of Improved Escherichia-Pseudomonas shuttle vectors
derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.";
RL Gene 148:81-86(1994).
DR EMBL: L30112; AAA6058.1; -.
DR HSSP: P0810; IXPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta_lactamase_1.
DR PRINTS: PR00118; BLACTMASEA.
KW Plasmid.
FT NON_TER
SO SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 9.1e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMATLTKLLTGELITLASROQLDWMENDKVGAPLRSALPAGFTIADSGSGGSGSG 219
DB 2 AAMATLTKLLTGELITLASROQLDWMENDKVGAPLRSALPAGFTIADSGSGSGSG 61

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 62 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 105

RESULT 2

Q52330 PRELIMINARY: PRT: 105 AA.
AC 052330;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Tm1 Bta protein (Fragment).
OS Escherichia coli.
OC Plasmid RK2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;

QY 160 AMATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 219
Db 2 AMATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 61

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 9.1e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 219
Db 2 AMATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 61

RESULT 3

Q8RTD8 PRELIMINARY: PRT: 145 AA.

AC 08RTD8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE SHV-5 enzyme (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 62 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 105

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 62 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 105

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 36.7%; Score 493; DB 2; Length 145;
Best Local Similarity 68.1%; Pred. No. 6e-34;
Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

QY 100 AATWSDMTANILTTTIGPKELTAPLHNGDHVTRIDRNEPELNEAIPNDEDTMPV 159
Db 7 AAXYSDNSANLTLATYGGPAGLTAFLRQIGDNTYRLDRMETELNEALPGDADDTTPA 66

QY 160 AMATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 219
Db 67 SAATTLRKLLTGEILLTASRQOLIDWMEADKVGPLRSALPAGWFIADKSGAGERSRG 126

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
Db 127 IYALLGPNKAKERIVYITTSQATMDERNROIAEIGASLIKHW 144

DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6988, and E/99 5-2;
 RA Subramaniam G., Navaratnam P.;
 RT "SHV gene sequences from ceftazidime-resistant Escherichia coli
 isolates from UMC, Malaysia.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327049; AAK07464.1; -;
 DR EMBL: AF327051; AAK07466.1; -;
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 138
 SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 5,1e-31;
 Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;
 Oy 111 NLLTTTGGPRELTAFLHNMGDHYTRLDRWEPELNEAIPNDRDTTPVAMATTLRLLT 170
 Db 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDARDTTTPASMAATTLRLLT 69
 Oy 171 GELLTLASROQLIDMWEADKYVAGPLRSALPAGWFIADKSGAGSGRGITTAAGDPGKP 230
 Db 70 SQRLSARSGROLQOMVDDRVAGPLIRSVLPAGWFIADKSGKRGARGLVALLGPNNKA 129
 Oy 231 SRIVVIY 237
 Db 130 ERIVVIY 136

RESULT 6
 ID O9AMA1 PRELIMINARY; PRT; 138 AA.
 AC O9AMA1;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1B/25;
 RA Subramaniam G., Navaratnam P.;
 RT "SHV gene sequences from ceftazidime-resistant Escherichia coli
 isolates from UMC, Malaysia.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327050; AAK07465.1; -;
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 138
 SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 5,1e-31;
 Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

Oy 111 NLLTTTGGPRELTAFLHNMGDHYTRLDRWEPELNEAIPNDRDTTPVAMATTLRLLT 170
 Db 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDARDTTTPASMAATTLRLLT 69
 Oy 171 GELLTLASROQLIDMWEADKYVAGPLRSALPAGWFIADKSGAGSGRGITTAAGDPGKP 230
 Db 70 SQRLSARSGROLQOMVDDRVAGPLIRSVLPAGWFIADKSGKRGARGLVALLGPNNKA 129
 Oy 231 SRIVVIY 237
 Db 130 ERIVVIY 136

RESULT 7
 ID O9AMA2 PRELIMINARY; PRT; 139 AA.
 AC O9AMA2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=918;
 RA Subramaniam G., Navaratnam P.;
 RT "SHV gene sequences from ceftazidime-resistant Escherichia coli
 isolates from UMC, Malaysia.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327048; AAK07463.1; -;
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 139
 SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 139;
 Best Local Similarity 67.7%; Pred. No. 5,1e-31;
 Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;
 Oy 111 NLLTTTGGPRELTAFLHNMGDHYTRLDRWEPELNEAIPNDRDTTPVAMATTLRLLT 170
 Db 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDARDTTTPASMAATTLRLLT 69
 Oy 171 GELLTLASROQLIDMWEADKYVAGPLRSALPAGWFIADKSGAGSGRGITTAAGDPGKP 230
 Db 70 SQRLSARSGROLQOMVDDRVAGPLIRSVLPAGWFIADKSGKRGARGLVALLGPNNKA 129
 Oy 231 SRIVVIY 237
 Db 130 ERIVVIY 136

RESULT 8
 ID O9AMA0 PRELIMINARY; PRT; 138 AA.
 AC O9AMA0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.


```

DR   InterPro: IPR000871; Beta_lactamase_A.
DR   PRINTS: PR00118; BLACTAMASEA.
FT   NON_TER 1
FT   NON_TER 139
SQ   SEQUENCE 139 AA; 15257 MW; F514247C882442AD CRC64;

Query Match
Best Local Similarity 66.9%; Score 453; DB 2; Length 139;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

QY 111 MLITTTGGPKELAFLLNMGDHYTRLDREPELNEAIPNDERDTTTPVAMATTLRKILT 170
DB 10 HLLATYGGPAGLAFALRQIDGNYTRDRMETELNEALPGARDTTTPASMAATLRKILT 69
QY 171 GELLITLASRQQLIDMNEADKYVAPLRLSALPAGMFIADKSGAGSGSGITAAIPDGKP 230
DB 70 SORLSARSORQLQMMVDYRVAGPLRSVLPAGMFIADKSGAGSGSGITAAIPDGKP 129
QY 231 SRIIVYI 237
DB 130 ERIIVYI 136

RESULT 12
ID 053553 PRELIMINARY; PRT; 67 AA.
AC 053553;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81098; AAB35839.2; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 67
SQ SEQUENCE 67 AA; 7287 MW; 85F8C85B0363F0CB CRC64;

Query Match
Best Local Similarity 25.1%; Score 338; DB 2; Length 67;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSRIDAGEQGLGRIRIHSQNDLVESPYTERHLDGMYVRELCSAITSNDTAAALL 114
DB 1 VLSRYVAGEQGLGRIRIHSQNDLVESPYTERHLDGMYVRELCSAITSNDTAAALL 60
QY 115 TTIGGPK 121
DB 61 TTIGGPK 67

RESULT 13
ID 08GDE5 PRELIMINARY; PRT; 95 AA.
AC 08GDE5;

```

```

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Ameyes S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
RT Chile."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY150585; AAN77730.1; -.
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 23.8%; Score 320; DB 2; Length 95;
Matches 61; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

QY 147 AIPNDERDTTTPVAMATTLRKILTGGELLTASRQQLIDMNEADKYVAPLRLSALPAGMFI 206
DB 1 ALPGDARDTTTPASMAATLRKILTSORLSARSORQLQMMVDYRVAGPLRSVLPAGMFI 60
QY 207 ADRSGAGSGSGITAAIPDGKPSRIIVYITGGS 241
DB 61 ADKTGASERGARGIYALLGPNNKAERIVYIYLRDS 95

RESULT 14
ID 09JN58 PRELIMINARY; PRT; 62 AA.
AC 09JN58;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81099; AAB35840.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 62
SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 23.5%; Score 316; DB 2; Length 62;
Matches 61; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 132 DHTVRLDWEPELNEAIPNDERDTTTPVAMATTLRKILTGGELLTASRQQLIDMNEADKY 191
DB 1 DHTVRLDWEPELNEAIPNDERDTTTPVAMATTLRKILTGGELLTASRQQLIDMNEADKY 60
QY 192 AG 193

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds

(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_37s_157t

Sequence: 1345
1 HPEITLVKVADEQICARVC.....TMDERNRQIAETIGASLIKHM 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	583	43.3	196	7	AAP60627
2	567	42.2	146	22	AAU23221
3	567	42.2	146	22	AAU23221
4	488	36.3	127	18	AAW20440
5	473	35.2	129	22	AAU23220
6	473	35.2	129	22	AAU23220
7	473	35.2	182	22	ABG27917
8	457	34.0	94	21	AAW59052
9	400	29.7	159	7	AAW60628

10	366	27.2	101	22	ABG27935	Novel human diagno
11	329	24.5	88	22	ABG27919	Novel human diagno
12	195	14.5	119	7	AAP60626	Beta-urogastrone -
13	145	10.8	51	21	AAV92783	Vlgs-beta-lactama
14	83	6.2	200	18	AAW55524	H. pylori ORF 29ep
15	80.5	6.0	158	22	AAU45138	Proionibacterium
16	75	5.6	202	24	AAE32763	Human zcyto24 prot
17	75	5.6	202	24	AAO16283	Mouse IMX129840-2
18	72.5	5.4	196	21	AAV73344	HTM clone 0258181
19	71	5.3	159	19	AAW69170	N-terminally tagged
20	70.5	5.2	193	23	AAU76854	Human integrin alp
21	70.5	5.2	193	23	AAU76854	Human integrin alp
22	70	5.2	15	10	AAP85803	Sequence encoded b
23	70	5.2	170	22	AAU52472	Proionibacterium
24	69.5	5.2	170	22	AAU56330	Proionibacterium
25	69.5	5.2	184	21	AAW25446	Pinus radiata cell
26	69	5.1	134	21	AAW27220	Zea mays protein f
27	69	5.1	166	23	AAW26537	Streptococcus poly
28	69	5.1	194	22	AAW81764	S. epidermidis ope
29	68	5.1	180	21	AAW20206	Arabidopsis thalia
30	67.5	5.0	152	22	ABG19138	Novel human diagno
31	67.5	5.0	152	22	ABG27771	Novel human diagno
32	67.5	5.0	187	23	AAU97104	Human MK61 protein
33	67	5.0	178	22	AAW91926	Human alpha 2 C4 a
34	67	5.0	202	22	AAW82686	Soybean calmodulin
35	66.5	4.9	83	22	AAU59092	Arabidopsis thalia
36	66.5	4.9	177	21	AAW41582	Neospora caninum a
37	66	4.9	183	19	AAW85905	Novel human diagno
38	66	4.9	202	24	AAE32764	
39	65.5	4.9	50	17	AAW03591	S. pneumoniae derl
40	65.5	4.9	136	21	AAV44229	Human zcyto25 prot
41	65	4.8	150	21	AAV95249	Human alpha 2 C4 a
42	65	4.8	168	21	AAW58698	Partial corn extra
43	65	4.8	203	18	AAW31265	Arabidopsis thalia
44	65	4.8	203	18	AAW31265	Arabidopsis thalia
45	65	4.8	207	22	ABG06157	Novel human diagno
					ABG19320	

ALIGNMENTS

RESULT 1	AAW60627	standard; Protein; 196 AA.
ID	AAW60627	standard; Protein; 196 AA.
XX	AAW60627	
AC	AAW60627	
XX	AAW60627	
DT	25-MAR-2003	(updated)
DT	17-JUN-1991	(first entry)
XX		
DE	Beta-urogastrone - beta-lactamase fusion protein from pUG2101.	
XX		
KW	Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;	
KW	fusion protein; beta-lactamase.	
XX		
OS	Synthetic.	
XX		
EH	Key	Location/Qualifiers
FT	Protein	1..120
FT		/label= beta-lactamase
FT	Misc-difference	121..123
FT		/label= adaptor
FT	Protein	124..196
FT		/label= beta-urogastrone
XX		
XX	DE3523634-A.	
PN		
XX	09-JAN-1986.	
PD		
XX	02-JUL-1985;	85DE-3523634.
PF		
XX	02-JUL-1984;	84UP-0137691.
PR		
XX		

PA	(EART) EARTH CHEM CO LTD.
XX	
P1	Aoki S., Ohgai H., Horinaka A., Hiramatsu H., Kounoto S., Nishimura A;
P1	Matsushiro S.;
XX	
DR	WPI; 1986-015031/03.
XX	
N-PSDB:	AAN60631.
PT	New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PF	and transformed cells contg. it.
XX	
PS	Disclosure: Page 56-59; 92pp; German.
XX	
CC	The fusion protein is less easily degraded by proteases and so
CC	protects beta-urogastrone and beta-lactamase collects in the periplasm
CC	of E.coli. It is therefore easy to collect and purify the product.
CC	Beta-urogastrone is the hormone of the salivary glands which suppresses
CC	stomach acid secretion and promotes cell growth, so is useful for
CC	treating ulcers and wounds. Previously the product was obt'd. only
CC	in small amts. from human urine.
CC	See also AAN60628. and 30-32.
CC	(Updated on 25-MAR-2003 to correct PA field.)
XX	
SQ	Sequence 196 AA:
Query Match	43.3%; Score 583; DB 7; Length 196;
Best Local Similarity	98.3%; Pred. No. 7.6e-52;
Matches 115;	Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY	1 HPETLVKKKDAEDDLGARVGIETLDLNSGKITLESFRSEPRPPMSTFVLLCGAVLSRID 60
Db	24 HPETLVKKKDAEDDLGARVGIETLDLNSGKITLESFRPEPRPMSTFVLLCGAVLSRVD 83
OY	61 AGQEQLGRRIHNSQNDLYEVSPTTEKHITDGMTVEELCSAAITMSDNTAANILLTTI 117
Db	84 AGQEQLGRRIHNSQNDLYEVSPTTEKHILTDGMTVEELCSAAITMSDNTAANILLTTI 140
RESUFT 2	
AAU23221	
ID	AAU23221 standard; Protein; 146 AA.
XX	
AC	AAU23221;
XX	
Dt	18-DEC-2001 (first entry)
XX	
DE	Novel human enzyme polypeptide #307.
XX	
KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW	ligase; hyperproliferative disorder; immunodeficiency disorder;
KW	autoimmune disorder; neurological disorder; metabolic disorder;
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW	blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW	nephrotropic; anticoagulant.
XX	
OS	Homo sapiens.
XX	
PN	WO200155301-A2.
XX	
PD	
PF	02-AUG-2001.
XX	
17-JAN-2001;	2001MO-US01239.
XX	
31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000;
PR	24-FEB-2000;
PR	02-MAR-2000;
PR	16-MAR-2000;
PR	17-MAR-2000;
PR	18-APR-2000;
PR	19-MAY-2000;
PR	07-JUN-2000;
PR	28-JUN-2000;

PR	07-JUN-2000	2000US-0215135
PR	07-JUL-2000	2000US-0216647
PR	07-JUL-2000	2000US-0216880
PR	11-JUL-2000	2000US-0217148
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225213
PR	14-AUG-2000	2000US-0225214
PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225269
PR	18-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	03-SEP-2000	2000US-0229505
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0233198
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0234001
PR	14-SEP-2000	2000US-0234063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234274
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236927
PR	29-SEP-2000	2000US-0236927
PR	29-SEP-2000	2000US-0236928
PR	29-SEP-2000	2000US-0236969
PR	29-SEP-2000	2000US-0236970
PR	02-OCT-2000	2000US-0236602
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	12-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239353
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241866

PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 11-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465566/50.
 DR N-PSDB; AAS41091.
 XX
 XX Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 XX diseases -
 PS Claim 11; SEQ ID No 1217; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the

CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAV22915-AAV23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 146 AA;
 Query Match 42.2%; Score 567; DB 22; Length 146;
 Best Local Similarity 97.4%; Pred. No. 2, 2e-50;
 Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 HPETLVKVDAREQDQGVYQLDLSNGKILSEFSRSEKRPFMSTFKVLCGAVSRID 60
 Db 28 HPETLVKVDAREQDQGVYQLDLSNGKILSEFSRSEKRPFMSTFKVLCGAVSRID 87
 QY 61 AGGEQLGRRIHYSQNDLVEYSPYTERKLRIDGKMTVRELCSAATMSDNTAAILLT 115
 Db 88 AGGEQLGRRIHYSQNDLVEYSPYTERKLRIDGKMTVRELCSAATMSDNTAAILLT 142
 RESULT 3
 ID AAM90870
 ID AAM90870 standard; Protein; 146 AA.
 AC AAM90870;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human Immune/haematopoietic antigen SEQ ID NO:18463.
 XX
 KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
 XX cytostatic; gene therapy; vaccine; metastasis.
 OS Homo sapiens.
 PN WO200157182-A2.
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.

PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225269
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225575
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	14-AUG-2000	2000US-0226279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0227182
PR	30-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232237
PR	14-SEP-2000	2000US-0232396
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234224
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236332
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246530.
PR 08-NOV-2000; 2000US-0246531.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK63651.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
PI

XX
XX
XX Claim 11: SEQ ID NO 18463; 3071pp + Sequence Listing; English.
PS
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (II)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the present invention.

SQ Sequence 146 AA; 42.2%; Score 567; DB 22; Length 146;
Query Match Best Local Similarity 97.4%; Pred. No. 2.2e-50;
Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 HPELVKVKDAEDQAGAVGYIELDNSGKILSFSESEPPMSTFVLLCGAVLSRID 60
|||||
DB 28 HPELVKVKDAEDQAGAVGYIELDNSGKILSFSESEPPMSTFVLLCGAVLSRID 87
|||||

QY 61 AGOEOLGRRIRHYSQNDLVEYSPVTEKHLTDGMTVRELXKAITMSDNTAANLLLT 115
|||||
DB 88 AGOEOLGRRIRHYSQNDLVEYSPVTEKHLTDGMTVRELXKAITMSDNTAANLLLT 142
|||||

RESULT 4
AAW20440
ID AAW20440 standard; protein; 127 AA.
XX
AC AAW20440;
XX
DT 14-JUN-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 34574062.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
KW replication; transcription; recombination; repair.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 121 /note= "encoded by codon YGC"
FT Misc-difference 122 /note= "encoded by codon WGC"
FT
XX
FN W09640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
PA (ASTR) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgaerd BL;
XX
DR WPI: 1997-052306/05.
DR N-PSDB; AAT67613.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
PS Claim 61; Pages 615-616; 1481pp; English.

CC The present sequence is a Helicobacter pylori cytoplasmic protein
CC involved in genomic replication, transcription, recombination and repair.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be

CC Isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 127 AA; 36.3%; Score 488; DB 18; Length 127;
Query Match Best Local Similarity 99.0%; Pred. No. 2.7e-42;
Matches 96; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPELVKVKDAEDQAGAVGYIELDNSGKILSFSESEPPMSTFVLLCGAVLSRID 60
|||||
DB 24 HPELVKVKDAEDQAGAVGYIELDNSGKILSFSESEPPMSTFVLLCGAVLSRID 83
|||||

QY 61 AGOEOLGRRIRHYSQNDLVEYSPVTEKHLTDGMTVREL 97
|||||
DB 84 AGOEOLGRRIRHYSQNDLVEYSPVTEKHLTDGMTVREL 120
|||||

RESULT 5
AAU23220
ID AAU23220 standard; protein; 129 AA.
XX
AC AAU23220;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #306.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytosolic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN W020015301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.

QY 1 HPETLVKVADEQLGARVGYIELDNLSKILESFSFEPRPMASFFKVLICGAVLSRID 60
Db 28 HPETLVKVADEQLGARVGYIELDNLSKILESFSFEPRPMASFFKVLICGAVLSRID 87
QY 61 AGOEOIGRRIRHYSQNDLVESPYTEKHLITDCMTVRE 96
Db 88 AGOEOIGRRIRHYSQNDLVESPYTEKHLITDCMTYXE 123

RESULT 6
AAM90871
ID AAM90871 standard; protein; 129 AA.
XX
AC AAM90871;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18464.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209667.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234423.
PR 21-SEP-2000; 2000US-0234424.
PR 25-SEP-2000; 2000US-0234497.
PR 25-SEP-2000; 2000US-0234498.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.


```

XX AAB59052;
AC
XX
XX 27-MAR-2001 (first entry)
XX
XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
DE
XX
XX Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX nontropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX MO200055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000MO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX MPI: 2000-611515/58.
XX
XX N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiant activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemia; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 94 AA;
XX
XX Query Match 34.0%; Score 457; DB 21; Length 94;
XX Best Local Similarity 97.8%; Pred. No. 2,8e-39;
XX Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

RESULT 9
AAP60628
ID AAP60628 standard; Protein; 159 AA.
XX
XX AAP60628;
AC
XX
XX 25-MAR-2003 (updated)
XX 17-JUN-1991 (first entry)
XX
XX Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX /label= beta-lactamase
XX MISC-difference 97..100
XX /label= adaptor
XX Protein 101..159
XX /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
XX Matsushiro S;
XX MPI; 1986-015031/03.
XX
XX N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX and transformed cells contg. it.
XX
XX Disclosure; Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E. coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obtd. only
XX in small amts. from human urine.
XX See also AAN60628, and 30-32.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 159 AA;
XX
XX Query Match 29.7%; Score 400; DB 7; Length 159;
XX Best Local Similarity 68.3%; Pred. No. 4.6e-33;
XX Matches 86; Conservative 8; Mismatches 20; Indels 12; Gaps 2;

```


Db 1 TLVKVDAEDQLGARVGYIELDNSGKILSFREPERFPMSTFKVLLCGAVLSRIDAQ 60
 QY 64 EQLGRR-----HYSNDIV 78
 DB 61 EQLGRRITILRMWTWLSHOSKSL 85

RESULT 12

AAp60626
 ID AAP60626 standard; Protein; 119 AA.

AC AAP60626;

DT 25-MAR-2003 (updated)

DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

KM Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 fusion protein; beta-lactamase.

OS Synthetic.

FT Key Location/Qualifiers

FT Protein 1..63

FT Misc-difference 64..66 /label= beta-lactamase

FT Protein 67..119 /label= adaptor

FT Protein /label= beta-urogastrone

DE3523634-A.

PD 09-JAN-1986.

PF 02-JUL-1985; 85DE-3523634.

PR 02-JUL-1984; 84JP-0137691.

PA (EART) EARTH CHEM CO LTD.

PI Aoki S, Ohgai H, Horioka A, Hiramatsu H, Koumoto S, Nishimura A;
 Matsushiro S;

DR WPI: 1986-015031/03.

DR N-PSDB; AAN60630.

PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 and transformed cells contg. it.

PS Disclosure: Page 55; 92pp; German.

CC The fusion protein is less easily degraded by proteases and so
 protects beta-urogastrone and beta-lactamase collects in the periplasm
 of E.coli. It is therefore easy to collect and purify the product.
 CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 stomach acid secretion and promotes cell growth, so is useful for
 CC treating ulcers and wounds. Previously the product was cold. only
 CC in small ants. from human urine.
 CC See also AAN60628, and 30-32.
 CC (Updated on 25-MAR-2003 to correct PA field.)

CC Sequence 119 AA;

Query Match 14.5%; Score 195; DB 7; Length 119;

Best Local Similarity 97.5%; Pred. No. 4.7e-12;

Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPETLVKVDADQLGARVGYIELDNSGKILSFSEER 40
 DB 24 HPETLVKVDADQLGARVGYIELDNSGKILSFSEER 63

RESULT 13
 AAY92783
 ID AAY92783 standard; Protein; 51 AA.

AC AAY92783;

DT 29-AUG-2000 (first entry)

DE Vtgs-beta-lactamase fusion protein (partial).

KM Vtgs; vitellogenin; secretory signal sequence; gene expression;
 oestrogen receptor binding protein; systemic circulation;
 beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

FT Key Location/Qualifiers

FT Peptide 1..21

FT Cleavage-site 15..16 /label= Vtgs

FT Protein 22..51 /note= "beta-lactamase mature protein"

W0200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WC-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

PA (VYSI-) UNIV SINGAPORE NAT.

PA (LAMT/) LAM T J.

PI Ding JL, Tan NS, Ho B;

DR WPI: 2000-365615/31.

DR N-PSDB; AAA28507.

PT Isolated nucleic acid for assaying for heterologous gene expression,
 detecting presence of compound that binds to estrogen receptor or
 producing desired protein from host cell comprises nucleotide sequence
 encoding secretory signal sequence

PS Example 6; Fig 14A; 73pp; English.

CC A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pBADVtgbLactKana was
 CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

CC Sequence 51 AA;

Query Match 10.8%; Score 145; DB 21; Length 51;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PETLVKVDADQLGARVGYIELDNSGK 30
 DB 23 PETLVKVDADQLGARVGYIELDNSGK 51

XX Sequence 158 AA;

Query Match 6.0%; Score 80.5; DB 22; Length 158;
Best Local Similarity 33.3%; Pred. No. 4.8;
Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

OY 110 ANLLITGGPEELAFPHNMGDHTRLDREPELENAIPNDRDITTVAAATIR--K 167
 DB 80 ARMATSTGCAPRVSSASLTARCAASWTYGPWEPSTRPAL-----TTSAPLAFRDITRSGR 134
 OY 168 LLTGELTLARQQLIDW 185
 DB 135 GRMGSWTLATRIPLIRW 152

Search completed: September 10, 2003, 12:24:52
Job time : 32.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds

(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_37s_157c

Sequence: 1345
1 HPEITAKVDAEDQAGARVG.....TMDENNRQIAETGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/3A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/3B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	74.5	5.5	124	US-08-311-731A-202	Sequence 202, App
2	72.5	5.4	197	US-09-252-991A-24241	Sequence 24241, A
3	71	5.3	159	US-08-991-890-4	Sequence 4, Appl1
4	69	5.1	169	US-09-328-352-4491	Sequence 4491, Ap
5	65.5	4.9	158	US-09-010-809-19	Sequence 19, Appl
6	65	4.8	150	US-09-239-909-2	Sequence 2, Appl1
7	64.5	4.8	178	US-09-252-991A-29942	Sequence 29942, A
8	64.5	4.8	198	US-09-413-814-87	Sequence 87, Appl
9	64	4.8	203	US-08-624-677A-2	Sequence 2, Appl1
10	63.5	4.7	144	US-08-225-480-4	Sequence 4, Appl1
11	63.5	4.7	144	US-09-118-445-4	Sequence 4, Appl1
12	63.5	4.7	160	US-09-252-991A-24737	Sequence 24737, A
13	63.5	4.7	170	US-08-858-207A-519	Sequence 519, App
14	62	4.6	132	US-09-252-991A-22681	Sequence 22681, A
15	62	4.6	181	US-08-482-142-195	Sequence 195, App
16	62	4.6	181	US-08-478-572-195	Sequence 195, App
17	62	4.6	181	US-08-484-296-195	Sequence 195, App
18	61.5	4.6	142	US-09-345-473E-31	Sequence 31, Appl
19	61	4.5	144	US-09-252-991A-55578	Sequence 25578, A
20	61	4.5	174	US-08-557-122A-6	Sequence 6, Appl1
21	61	4.5	174	US-09-262-666-6	Sequence 6, Appl1
22	61	4.5	189	US-08-671-548C-48	Sequence 48, Appl
23	61	4.5	200	US-08-557-122A-12	Sequence 12, Appl
24	61	4.5	200	US-09-262-666-12	Sequence 12, Appl
25	60.5	4.5	103	US-09-733-210-1282	Sequence 1282, Ap
26	60.5	4.5	167	US-08-690-849-2	Sequence 2, Appl1
27	60.5	4.5	167	US-09-004-053-2	Sequence 2, Appl1

28	60	4.5	145	US-09-134-001C-4982	Sequence 4982, Ap
29	59.5	4.4	146	US-09-134-001C-5269	Sequence 5269, Ap
30	59	4.4	171	US-09-107-532A-3579	Sequence 3579, Ap
31	59	4.4	180	US-09-194-905-5	Sequence 5, Appl1
32	59	4.4	193	US-09-252-991A-30363	Sequence 30363, A
33	58.5	4.3	203	US-09-252-991A-24921	Sequence 24921, A
34	58	4.3	136	US-09-252-991A-30622	Sequence 30622, A
35	58	4.3	168	US-08-451-947-6	Sequence 6, Appl1
36	58	4.3	168	US-08-424-826A-6	Sequence 6, Appl1
37	58	4.3	168	US-08-928-694-6	Sequence 6, Appl1
38	58	4.3	168	US-08-450-842-6	Sequence 6, Appl1
39	58	4.3	168	US-08-451-390-6	Sequence 6, Appl1
40	58	4.3	168	PCT-US91-06950-6	Sequence 6, Appl1
41	58	4.3	183	US-09-252-991A-31347	Sequence 31347, A
42	58	4.3	187	US-09-252-991A-29325	Sequence 29325, A
43	58	4.3	189	US-09-252-991A-17160	Sequence 17160, A
44	58	4.3	204	US-09-252-991A-17611	Sequence 17611, A
45	58	4.3	206	US-09-029-603-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311, 731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202
Query Match 5.5%; Score 74.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;
Oy 130 MGDHV--TRIDRMEPELN---EALPNDEROT---TTPVAMATTLRKLTLGELLTLASRO 160

[illegible]

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,8990
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-991-890-4

Query Match          5.38; Score 71; DB 3; Length 159;
Best Local Similarity 23.88; Pred. No. 6.5;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

49 VILGAVLSRIDGOE-----QLGRRIHYSQN-----DLVEKSPYTEKH 87
   ||||| : | | | | | : | | | | | : | | | | | : | | | | |
Db      12 LLIGGNV--VSPQELHAERGGRRHHNHNHGSSGAELGCCPRFGKIHLSCYCPMEKT 69
Oy      88 LTDCMYVELCSAAITMSDNTAAILTTTGPRELTAFLHMNGDHTRLDREPELNKA 147
       70 FT-----TTPGGWLES-GRPKEVSTSNRDKQAL-----GTTSEP 105
Oy      148 IPNDERDTTPPVAMA-TTLRKLL 169
       || | | : | : | : | : | : | : | : | : | : | : | : |
Db      106 IPNLSPELKRPESGPSIKKKI 128

RESULT 4
US-09-328-352-4491
Sequence 4491, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Berton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4491
LENGTH: 169
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

Query Match          5.1%; Score 69; DB 4; Length 169;
Best Local Similarity 22.2%, Pred. No. 12;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

118 GGPRLTAFAFLNMHDGYTRLDRWEPELMSEAIPNERDPT-----TPVMATTLKRLL 169
    | | | | : | | | : | | | : | | | : | | | : | | | : | | | : |
Oy      5 GEGTIMTKMLAQDN---RFDMWEIQYL---DLTVREIRIMVVTTIFIVVIIVIGS.58
Db      170 TGEIITLASROO-----LIWMEDKVAGPLLRSALPWGFIAIDKSGAGER--GSRG 219
           ::::|::|         ::::::|         ||| ||| : | : | : |
Db      59 LFKMHSLEAOOKRLNDLKDLMVMMOSNAV-----TWKPANELEDKSKGIQHVAODOG 112
               ::::|::|         ::::::|         ||| ||| : | : | : |
Oy      220 ITAALGPDCKPSRYIVYTTSQAETMDERNQAIENIGSLK 261

```


SEQ ID NO 87
LENGTH: 198
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match 4.8%; Score 64.5; DB 3; Length 198;
Best Local Similarity 22.4%; Pred. No. 48;
Matches 62; Conservative 30; Mismatches 78; Indels 107; Gaps 15;

QY 1 HPELVVKKAEQDLGARVGIETDLNSKILSFSEERFPMSTP-----KYLGCAY 55
DB 7 NPEAVDVKKKTS--AAFGV---DAIVHVKLEIGAGVETALVTGEGHEPESVLCVAS 60
QY 56 LSRIDAGOEQGRRIHVSQNDLVESPYTEKHLDGTVRELCSAATMSDNFTANLLT 115
DB 61 LVR-----MGVRIAAA---IDROAD--ILR 82
QY 116 TIGPKRELTAFLNNMGDHYTRLDMPELNAIPNDERDPTTPVAMATTLKLLTGELLT 175
DB 83 AVG-----ATRVQLTEEMGRV---GADITMPLAQ-----DLDD 114
QY 176 LASRQOLDMKADKAVCPILRSALPAGWFIADK---SGAGER-----GSRGIIALG 225
DB 115 LASHYRVVPW---NAHGPIVGQTL-AGSKTRQRYRINVLGVRHTRKRPQDKRLEAPT 169
QY 226 PDGKPRIVVIYTTGSQATM-----DERNRQIAEIG 256
DB 170 PD-----YVIRDDGTLLVGDSDVSFVAEIG 197

RESULT 9

US-08-624-677A-2
Sequence 2, Application US/08624677A
Patent No. 6476192

GENERAL INFORMATION:

APPLICANT: Lally, Nicola C.
APPLICANT: Jenkins, Mark C.
APPLICANT: Dubey, Jitender P.
TITLE OF INVENTION: Antigens Useful for the Serodiagnosis of
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Graeter, Janelle S.
STREET: Rm. 411, Bldg. 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,677A
FILING DATE: 15-Apr-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0228,95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-6629
TELEFAX: 301-504-5060

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-677A-2

Query Match 4.8%; Score 64; DB 4; Length 203;
Best Local Similarity 23.2%; Pred. No. 56;
Matches 38; Conservative 22; Mismatches 74; Indels 30; Gaps 6;

QY 54 AVLSRIDAGOEQGRRIHVSQNDLVESPYTEKHLDGTVRELCSAA--TMSDNTAAN 111
DB 33 AGVSNTDGGDDDAAGNPVD-----SDTDAITDGEWPRVSGQKHTYOKGSLIK 83
QY 112 LLLTTIGPKRELTAFLNNMGDHY-TRLDHPELNAIPNDERDPT-----T 157
DB 84 LAVPVGA---LTSYL--VADRVLPBLTSAEBSGTSISGKKRVKAVGAIALVMAAARA 138
QY 158 PVAMATTLKRLTGELLTLASRQOLDMKADKAVCPILRSALP 201
DB 139 GLGLAFTFRHFVKKSKTVASEDSALGNSEQYVEGTVNGSSDP 182

RESULT 10

US-08-225-480-4
Sequence 4, Application US/08225480
Patent No. 5817793

GENERAL INFORMATION:

APPLICANT: LEVY, STUART B.
TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,085
FILING DATE: 28-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: T0359/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-225-480-4

Query Match 4.7%; Score 63.5; DB 2; Length 144;
Best Local Similarity 29.5%; Pred. No. 38;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRRIHY--SQND--LVESY-SP-VTEKHLDGTVRELCSAATMSDNFTANLLTTIG 118
DB 14 LGRLIHVNQKKRDLNLEYSPLDITAAQFVLCSIR--CAACIT----- 56
QY 119 GKPELTAFLA-NMGDHYTRLDR-----WPEPLNAIPNDERD-----TTTPVAMATTLRK 167
DB 57 -PVELKRVLSVDIGALTRMDRLVCKGWERTLPN--PNDRKGVLYKLTGGAALCEQCHQ 113
QY 168 LLTGELLTLASRQOLDMKADKVA--GPIILRSALP 201

Db 114 LVGGDL-----HQLTKNLTADVATLEYLKKVLP 144

RESULT 11
US-09-118-445-4
; Sequence 4, Application US/09118445
; Patent No. 6391545
; GENERAL INFORMATION:
; APPLICANT: LEVY, STUART B.
; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
; ASSAYS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,445
; FILING DATE: 17-Jul-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,480
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: 70359/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-118-445-4

Query Match
Best Local Similarity 4.7%; Score 63.5; DB 4; Length 144;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

Db 66 LGRRIHY--SOND--LVEY--SP--VTEKRLDGMVRELCSAITSNDTANLLITTTG 118
14 LGRLLHMVOKKDRLENTLSPLDITMAQFKVLCISIR--CAACIT----- 56

Db 119 GKELTAFLH-NMGDHTVRLDR-----WPELNEALPDERD-----TTTPVAMATTLR 167
57 -EVELKKVLSVDLALTRMLDLVCKGWERLPN--PNKRGVLYKLTIGALICEQCHQ 113

Db 168 LITGELLTASRQQLIDMMEADKVA--GPLRSALP 201
114 LVGGDL-----HQLTKNLTADVATLEYLKKVLP 144

RESULT 12
US-09-252-991A-24737
; Sequence 24737, Application US/09252991A
; Patent No. 6351795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24737
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24737

Query Match
Best Local Similarity 4.7%; Score 63.5; DB 4; Length 160;
Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

Db 3 ETLVKVDAEDQIGARVGYELDINSKTL-----ESFSREERFPMS 45
109 EILEKVSFLKSGGAIVHLEVDENPNKAVSFYKSGFERERFVMS 156

RESULT 13
US-08-858-207A-519
; Sequence 519, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimm, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 519:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
US-08-858-207A-519

Query Match
Best Local Similarity 4.7%; Score 63.5; DB 4; Length 170;
Matches 21.1%; Pred. No. 49;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_37S_157T

Sequence: 1 HETLVKVDADQAGARVG.....TMDERNKQIAETGASLIKHW 263

Scoring table: BLOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	457	34.0	94	US-10-102-806-760	Sequence 760, App
2	75	5.6	202	US-10-127-816-9	Sequence 9, Appl
3	75	5.6	202	US-10-142-717-12	Sequence 12, Appl
4	70.5	5.2	193	US-09-805-354-8	Sequence 8, Appl
5	70.5	5.2	193	US-10-144-259-8	Sequence 8, Appl
6	70	5.2	206	US-10-156-761-13867	Sequence 13867, A
7	69.5	5.2	184	US-10-101-464A-765	Sequence 765, App
8	69	5.1	149	US-10-156-761-8136	Sequence 8136, App
9	67.5	5.0	149	US-10-233-926-4	Sequence 4, Appl
10	67.5	5.0	187	US-09-948-018-8	Sequence 8, Appl
11	67	5.0	178	US-09-738-626-5680	Sequence 5680, App
12	67	5.0	194	US-10-156-761-8623	Sequence 8623, App
13	66	4.9	202	US-10-127-816-11	Sequence 11, Appl
14	64.5	4.8	189	US-09-815-242-5862	Sequence 5862, App
15	64.5	4.8	189	US-09-815-242-12979	Sequence 12979, A

16	64.5	4.8	191	US-10-156-761-8978	Sequence 8978, App
17	64	4.8	162	US-09-738-626-4796	Sequence 4796, App
18	64	4.8	179	US-09-764-868-757	Sequence 757, App
19	64	4.8	179	US-10-106-698-4858	Sequence 4858, App
20	64	4.8	15	US-10-189-346-12	Sequence 12, Appl
21	63.5	4.7	88	US-10-100-252-6	Sequence 6, Appl
22	63.5	4.7	144	US-10-131-406-4	Sequence 4, Appl
23	63.5	4.7	202	US-10-156-761-13460	Sequence 13460, A
24	63.5	4.7	206	US-10-738-626-5425	Sequence 5425, App
25	62.5	4.6	160	US-09-882-227-230	Sequence 230, App
26	62.5	4.6	190	US-10-156-761-9507	Sequence 9507, App
27	62.5	4.6	191	US-09-764-870-390	Sequence 390, App
28	62.5	4.6	174	US-10-125-540-390	Sequence 35777, A
29	62	4.6	9	US-09-864-761-35777	Sequence 523, App
30	62	4.6	189	US-10-125-540-523	Sequence 523, App
31	62	4.6	191	US-09-927-827-64	Sequence 64, Appl
32	62	4.6	11	US-09-862-027-31	Sequence 31, Appl
33	61.5	4.6	196	US-09-738-626-6624	Sequence 6624, App
34	61.5	4.6	185	US-10-101-464A-616	Sequence 616, App
35	61	4.5	202	US-10-189-346-16	Sequence 16, Appl
36	60.5	4.5	190	US-09-738-626-4637	Sequence 4637, App
37	60	4.5	182	US-10-136-761-12056	Sequence 12056, A
38	59.5	4.4	196	US-09-741-669-461	Sequence 265, App
39	59.5	4.4	206	US-09-815-242-10335	Sequence 461, App
40	59.5	4.4	9	US-09-867-550-1868	Sequence 10335, A
41	59.5	4.4	111	US-09-768-225B-40	Sequence 1860, App
42	59	4.4	134	US-09-922-683-5	Sequence 40, Appl
43	59	4.4	180	US-10-156-761-10043	Sequence 5, Appl
44	59	4.4	15		Sequence 10043, A
45	59	4.4	181		

ALIGNMENTS

```
RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P4103PICI
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-760
```

Query Match 34.0%: Score 457; DB 15; Length 94;
Best Local Similarity 97.8%: Pred. No. 1.1e-39;

Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
0Y 75 NDLYEVSPTKEHLIDGTFVREICSAATMSDNTAANLLTTIGGPKRLTFALNMGDRV 134
DB 1 NDLYEVSPTKEHLIDGTFVREICSAATMSDNTAANLLTTIGGPKRLTFALNMGDRV 60
```



```

US-10-144-259-8
; Sequence 8, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amln
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-348001
; CURRENT APPLICATION NUMBER: US/10/144, 259
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8

```

```

Query Match
Best Local Similarity 5.2%; Score 70.5; DB 15; Length 193;
Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

QY 59 IDAGQDELGRIRHNSQNDLYEISB--VTEKHLTDGKMTRELCSAATMSDNTAANLLT 116
Db 36 IGPQQLQVG-----VVQGEDVYHEFLNDYSVKDYVAASHIEGR----- 77

QY 117 IGGPKELAFPLHNGDHTRLDRNE-----PELNEALPDERD 154
Db 78 -GGTEFRITAF-----GIEFARSEAFQKGRGAKKMAIVITDGSHPDLERVIQOSERD 132

QY 155 TTPPVAMA 162
Db 133 NVTRYAVA 140

```

```

RESULT 6
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMCRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13867
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867

```

```

Query Match
Best Local Similarity 5.2%; Score 70; DB 15; Length 206;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;

QY 70 IHSQNDLYEISVTEKHLTDGKMTV-----RELCSAATMSDNTAANLLTIT--G 118

```

```

Db 1 VHASOG---AQGPOTESAGEKYTMQAAQGESPREFA-----NRSSALLKTVANLSG 50
QY 119 GPKR-----LTALPHNGDHTVTRLD-----RW-----EPLENEAIPNDE 152
Db 51 GDRHMAEDLLQNLKRAEDWRSRIDEPAVYQVLYRQOVSFRRLKRRRELSVAP-PE 109
QY 153 RDTTPPVAMATYTRKLTGELTLASRQOL-----DMNEADKXAGPLLSALPAGNF 205
Db 110 ASTGPDAASAEELRLVWKGALRLATRNQRTVLRYFEDELPEAD--VARLIGSVGTRS 167
QY 206 IADKSGAGERGSRGIIAALGP---DGKPSR 232
Db 168 TTRSLARLETLAPELAAIGPADAEQEPSR 197

```

```

RESULT 7
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101, 464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; NUMBER OF SEQ ID NOS: 2000-01-11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-765

```

```

Query Match
Best Local Similarity 5.2%; Score 69.5; DB 15; Length 184;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;

QY 41 PPMSTFRVYL-----LCGAVLSRIDAGQELGRIRHNSQNDLYEISPV-----TERKHLTDG 91
Db 23 FCMSTERILVYPYMLNGSVASRL-----RDSINGKPALDMPTRKRRLSLG 67
QY 92 -----MTVRELCSAATMSDNTAANLL-----TTIGGKELTAFPLHNGDHTVTRLDN- 140
Db 68 AARGLIYIHEQCQPIIHRDVKAAANILDERFEAVVGGFGAKLDRHDSHTTAVRGTV 127
QY 141 ---EPLENEAIPNDEDTTPPVAMATYTRKLTGEL---LTIASRQO--LIDDM 186
Db 128 GHIAPEYISTGGSSEK--TDVFGFGLLELTITGKALDPFGAANQKGVMLDW 179

```

```

RESULT 8
US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMCRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA

```

```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8136.

Query Match
Best Local Similarity 5.18; Score 69; DB 15; Length 149;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

QY 168 LITGELTL-----ASROOLIDMEADKVAAGPLRSALPAGWFIADKSG----- 211
DB 50 LLDIELTLIKRLVLVASYDAKAKEMIDWEND-----FALSHADCGRELAENRRLREI 105
QY 212 AGERSRGITIALGPDGKPSR 232
DB 106 AGLRDAQALPSAEGPAEKPER 126

RESULT 9
US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US20030131382A1
; GENERAL INFORMATION:
; APPLICANT: Allied, Stephen M.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLYLTRANSFERASE
; FILE REFERENCE: BRL419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
US-10-233-926-4

Query Match
Best Local Similarity 5.0%; Score 67.5; DB 16; Length 149;
Matches 22; Conservative 23; Mismatches 33; Indels 25; Gaps 4;

QY 159 VAMATTLRKLL-----TCELLTLASROOLIDMEADKVAAGPLRSALPAGWFIADKSGA 212
DB 14 LLSLSLSPFLMLMAADHAAEAAPQSOEEEDPKFAEGGDGY-----EVAIDRGG 64
QY 213 GERGSGITIALGPDGKPSRIV-----IYTTGSGATDERRNR 250
DB 65 GGAANGCI-----PGGRPIRYVADGIYDLFHFCHAKSLSEQAKR 102

RESULT 10
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US2002015097A1
; GENERAL INFORMATION:
; APPLICANT: Theill et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
```

```

; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-018-8

Query Match
Best Local Similarity 5.0%; Score 67.5; DB 10; Length 187;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

QY 112 LLITTTGGPKELFAFLNMGDHYTRIDRWEPE-----LNEALPDERDOTTTPVAMAT 163
DB 11 LLILALAPPEAS-----QYGRLEYWNPDKKCCSSCLQRRGP-----PPCPGALETG 58
QY 164 -----TLRLTNGELTLASR--QOLIDMEA-----DKVAGP-----L 195
DB 59 DTKREASLPLLSRELSSLASQPLSLDLELYLEELIYLDPEPPGGMANGTRHLA 118
QY 196 LRSALPAGWFIADKSGAGERSRGITIALGPDGKPSR--IVYITTSQATDERRRQIA 253
DB 119 ARYGLPAAW-----STFAVSLPSSRPLALIEWVAREPSSASLQICTHIA 165
QY 254 EIG 256
DB 166 QLG 168

RESULT 11
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US2002019605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680

Query Match
Best Local Similarity 5.0%; Score 67; DB 10; Length 178;
Matches 25; Conservative 17; Mismatches 37; Indels 22; Gaps 4;

QY 171 GELTLASROOLIDW-----MEADKVAAGPLRSALPAGWFIADKSGAGERSRGITIA 223
DB 111 | : : : : : | | | | | | | | | | | | | | | | | | | | | | |
```

Db 61 GELFDAEQRGELNFGAGTLEVTSGVDNPL---TLPRM-----RRNRGLVA 107
QY 224 LGPDGKP--SRIVYITTSQATMDERNRQIAEIGASLIRH 262
Db 108 LDQDGKKRVARIGALNDAEFHVLIERNKILLETTELH 148

RESULT 12
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYMUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match 5.0%; Score 67; DB 15; Length 194;
Best Local Similarity 38.6%; Pred. No. 80;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

QY 202 AGWFINDKSGAGERSRG-----IIALGPDG--KPSRIYVITTT 239
Db 75 SGRRTVPRSGAGERTEGAEFRITTAIGPLAVHEPVVAVVVT 118

RESULT 13
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match 4.9%; Score 66; DB 15; Length 202;
Best Local Similarity 21.5%; Pred. No. 1,1e+02;
Matches 41; Conservative 23; Mismatches 65; Indels 63; Gaps 8;

QY 50 LLOGAVLSRIDAGQEQGLRRIRHSND--LVEXSPYTERHL-----TDGMTVREL----- 97
Db 17 LLLAAVLTHTQADPPVPVPRATRLPVEAKDCHIAQFKSISPRELQAFKAKAIEKRLLEKDM 76

QY 98 -CSA-----AIIIMS--DNTAAILLTTTGGPRELTAFLN 129
Db 77 RCSHLSIRAMPDKQLOVERKALQAEVALLTKWENINDSALTITLIGOPHLTSLHS 136

QY 130 MGDHTV-----RLDRWEPELNEA---IPNDERDTTPVMAATLRLTGE 172
Db 137 QLOQTQLOATAPRPSPRSLRWLRLQEAQSKTPGCLSDSVT-----SWLPOLLARD 191

QY 173 LLLIASROOLI 183
Db 192 LKCVASGDQCV 202

RESULT 14
US-09-815-242-5862
; Sequence 5862, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5862
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5862

Query Match 4.8%; Score 64.5; DB 9; Length 189;
Best Local Similarity 20.5%; Pred. No. 1.4e+02;
Matches 42; Conservative 32; Mismatches 84; Indels 47; Gaps 8;

QY 4 TLVAVKDAEQDGLARVGYIELDLSNGKLESFRSEERFPMMSTFKVLGCA----- 54
Db 10 TLIRI--VQEDQNRFTFTELTKNIQDVVEMLNQO-----IEKLCUTGNGAVIAENT 61

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_37s_157t
Perfect score: 1345
Sequence: 1 HETLVKVKDAEDQIGARVG.....TMDERNROIAETGASLIRHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	524	39.0	105	2	JC2566
2	77	5.7	191	2	S67447
3	74.5	5.5	113	2	T45195
4	74.5	5.5	172	2	AD3606
5	72.5	5.4	184	2	T21126
6	72.5	5.4	200	2	G97064
7	71.5	5.3	152	2	T36984
8	71.5	5.3	195	2	T36975
9	71	5.3	153	2	E81708
10	71	5.3	177	2	A83753
11	71	5.3	192	2	G83096
12	70	5.2	108	2	E84217
13	70	5.2	170	2	AB0192
14	69.5	5.2	167	2	D87360
15	68.5	5.1	145	2	F84251
16	68.5	5.1	190	2	T09136
17	68.5	5.1	192	2	A83587
18	68	5.1	149	2	F71252
19	67.5	5.0	131	2	AD2861
20	67	5.0	180	2	C71885
21	66.5	4.9	116	2	C82906
22	66.5	4.9	177	2	D90227
23	66.5	4.9	198	2	D95285
24	65.5	4.9	181	2	F56978
25	65.5	4.9	181	2	AB1902
26	65.5	4.9	195	2	AE0623
27	65.5	4.9	197	2	H90211
28	65	4.8	148	2	E75283
29	65	4.8	150	2	T08585

30	65	4.8	160	1	E69186	conserved hypothet
31	65	4.8	168	2	B75498	conserved hypothet
32	65	4.8	177	1	B43387	nonstructural prot
33	65	4.8	177	2	J01931	nonstructural prot
34	65	4.8	180	2	G70912	hypothetical prote
35	64.5	4.8	145	2	H75062	hypothetical prote
36	64.5	4.8	151	2	D81333	probable protein-t
37	64	4.8	113	2	D70580	hypothetical prote
38	64	4.8	142	2	A71000	hypothetical prote
39	64	4.8	161	2	C71409	hypothetical prote
40	64	4.8	162	2	AG0769	probable RNA polym
41	64	4.8	177	2	T48420	probable acetyltra
42	64	4.8	178	2	I40124	hypothetical prote
43	64	4.8	191	2	E95333	outer surface prot
44	63.5	4.7	164	2	H82336	hypothetical prote
45	63.5	4.7	166	2	C90029	regulator of sigma

ALIGNMENTS

RESULT 1

JC2566
b1a protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C:Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148(128), 81-86, 1994

A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MES>

C:Genetics:

A:Gene: b1a

C:Superfamily: beta-lactamase I

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 7.6e-37;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	160	AMATTLRRLTGLGELLTLASROQLDMEADRVAGPLRSALPAGFIADKSGAGSGSG	219
DB	2	AMATTLRRLTGLGELLTLASROQLDMEADRVAGPLRSALPAGFIADKSGAGSGSG	61
QY	220	IIAALGPDKPSRIVVITTSQATMDERNROIAETGASLIRHW	263
DB	62	IIAALGPDKPSRIVVITTSQATMDERNROIAETGASLIRHW	105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996

A:Reference number: T21766

A:Accession: T38062

A:Status: Preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:Z69944; NID:q1217974; PIDN:CAA93808.1; PID:q1217978; GSPDB:C

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.7%; Score 77; DB 2; Length 191;

	Heat	Local similarity	22.6%	Pred.	No. 38:
	Matches	43;	Conservative	22;	Mismatches 69; Indels 56; Gaps 7,
OY	104	MSDNTAANLITTTGGPRELTAFLLHNNDHYTRLDRMEPELNALPDERDTTYYAMAT	163		
Dd	1	MSVHSHNLMQNSGIDKTAIALIVN---VAARD-----PASSKSTQAOLVSMLN	46		
OY	164	TLRKLITGELLTLTASRQOILDMWEADRYACPLRLSAPAGFIADKSAGAGENGSIITAA	223		
Dd	47	EFRCL--IURLPGLYKLIVNFRDSSPEYTSNAINIGGYVE--GLAFVGKROIISI	99		
OY	224	LGPDKP-----SRIVVVITYGSQATDEBNRRQI-----	252		
Dd	100	----SKPLEDKLMTLMSRFMLIDLTLTYQLLRRENTEDBEKEHQDLASNLASLPICIRHS	155		
OY	253	AEGISASLIKH 262			
Dd	156	VENGCAGLAKH 165			

```

RESULT 3
T45195
hypothetical protein u1756t [imported] - Mycobacterium leprae
C|Species: Mycobacterium leprae
C|Dates: 21-Jan-2000 #sequence_Revision 21-Jan-2000 #text_change 21-Jan-2000
C|Accession: T45195
R|Robison, K.
submitted to the EMBL Data Library, September 1994
A|Reference number: Z16911
A|Accession: T45195
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-113 <KEI>
A|Cross-references: EMBL:U15180; PIDN:AAA62885.1

```

	Query Match	5.5%	Score 74.5;	DB 2;	length 113;	
	Best Local Similarity	28.6%;	Pred. No. 30;			
	Matches	34;	Conservative 19;	Mismatches 43;	Indels 23;	Gaps 8;
Oy	130 MGDHV--TRLDWPEELN---EATPNDERD-----TTPVAMATLRKLITGCELTIASRQ	180				
	: :	: :	:	:	:	:
Dd	1 MGDDIGMREGRMTGNTOCPRLVPVGDESPILDGRASPEDLIIT--MLSLPTIMSHPPS	58				
Oy	181 QLIDIMEEA-DKVAGPLL-----RSALPWGFIAKDSGAGEGSGRCIIAALGPCKPFR	232				
Dd	59 RDDDWVEEPEDLQGVAFVDATGATMPAVGCI---GASTRGSS-GILATLSPTPRQPAR	112				
	: :	:: :	:	:	:	:

RESULT 4
 AD3606
 molybdopterin biosynthesis mog protein [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AD3606
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
 E.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leless
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference numbers: AD3252; PMID:11756688
 A:Accession: AD3606
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-172 <KUP>
 A:Cross-references: GB:AS008918; PIDN:AAL54015.1; PID:g17984966; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BME110773
 A:Map position: II
 C:Superfamily: molybdenum cofactor biosynthesis protein B moab

 Query Match 5.5%; Score 74.5; DB 2; Length 172;
 Best Local Similarity 25.5%; Pred. No. 53;
 Matches 40; Conservative 14; Mismatches 42; Indels 61; Gaps 9;

```

QY      83 VTEKHLTDGA--YVRELCSALITMSDNTAANLLTTIG--GPRELTAFLHNQDHYTRLDW 140
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      39 IVRVRIPIDGMESEYRD---TLIDLCPTVACDILLTTGTGSPS----- 76
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      141 EPELNEIAIPDEEDDTTPAAMATTLRLKLT---GELLTLASROOLIDMHEADKVAGPLRS 198
           || || || || || || || || || || || || || || || || || || || ||
Db      77 -----PRDE---TPPAMKAVLHAKELPGFGEGMRVSLSEQ----- 107
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      199 ALPAGWFIADKSGAGERSRGITIALGPDGKRSRIIV 235
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      108 -TPPA--VLSROTAGSRGKSFT---LNLPGKASTIAM 138
           : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
T21126
ADP-ribosylation factor homolog F19H8.3 [similarity] - *Caenorhabditis elegans*
C1:Species: *Caenorhabditis elegans*
C1:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
C1:Accession: T21126
R:Steward, C.
submitted to the EMBL Data Library, March 1997
A:Reference number: 219379

A:Accession: T21126
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-184 <M11>
A:Cross-references: EMBL,T293378; PTDN:CAD07583.1; GSPDB:GM00020; CESP:F19H8.3
A:Experimental source: Clone F19H8
C:Genetics:
A:Gene: CESP:F19H8.3
A:Map position: 2
A:Introns: 143/2
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:90-95/Region: nucleotide-binding motif B
F:112-113/Region: GTP-binding NKXD motif
F:2/Modified site: myristylated amino end (gly) (in mature form) *status predicted

	Query Match	5.48;	Score 72.5;	DB 2;	Length 184;	
	Best Local Similarity	21.68;	Pred. NO. 86;			
	Matches	46;	Conservative	38;	Mismatches	80; Indels 49; Gaps 11
QY	31 ILSEFSRSEERPFMMSTFKVLLCGAVLSRIDAGOEOLGRRIRHYSQNDLVESPYTEKHLTD	90				
	: : : : : :	:	:	:	:	:
Dd	6 VLKSKS-----PSGREIRILLGL-----DNAGKTTLIKQL--SSEDYQHATP-----TK	49				
QY	91 GMTVRELCSAATMSDNTANLILTTIGCPKEITALHNMGHV-----TRLD	138				
	: : : : : :	:	:	:	:	:
Dd	50 GNVTIVAA---MGD---TRLNWDMDGGORSLRPWSWYENIDTLLIVIDSNDKKRDD	102				
QY	139 RHEPELTNEAIIPNDERDTTPVAMATTLRKLITGELLTSRLOOLDIMEADRVAGPLRS	198				
	: : : : : :	:	:	:	:	:
Dd	103 EKNIELGELL-DEKRIKRVPVLIFPNKKO-----DLVTAASSEITRKMLND-----LLMD	151				
QY	199 ALPAGNFIADKSGAGERGSRGIITAAIGPPGKPS	231				
	: : : : : :	:	:	:	:	:
Dd	152 RT---WHIQACALKNEGINDGITWASWIKRA	181				

RESULT 6
G97064
spore coat protein CORC [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Rollins, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:g15024270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

Query Match 5.4%; Score 72.5; DB 2; Length 200;
Best Local Similarity 25.0%; Pred. No. 96;
Matches 35; Conservative 27; Mismatches 55; Indels 23; Gaps 7;

QY 1 HPELVKDAEQLGARVGIETLNSKILSEFR-SEERFPMSTFKYLLCGAVLSRI 59
DB 10 HP---VKIKPNQOLAKVITITQYGPDLASIRYLSQFSMTVP-----QALATLN 59
QY 60 DACQEQGRRIHYSQNDLVEYSPVTEKHLTDGTVRELCSAATF--MSDNTAANLLITTT 117
DB 60 DIGTEELAH-----LEIVGSIVRQLSRGLSVLELKSKSGLDVAFADHDSALYRASAA 110
QY 118 GGPKELEAFILNMGDHVTSL 137
DB 111 GNP-ETAAVYIQSKDPTTDL 129

RESULT 7

hypothetical protein SCJ11.13 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000

C:Accession: T36984

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999

A:Reference number: Z21618

A:Accession: T36984

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-152 <OLI>

A:Cross-references: EMBL:AL109949; PIDN:CA52898.1; GSPDB:GN00070; SCOEDB:SCJ11.13

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCJ11.13

C:Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
Best Local Similarity 23.4%; Pred. No. 80;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

QY 106 DMTANLLTTTIGGPELTAFLNMGDHYTRLDREPELNEAIPNDERDTPVAMATTL 165
DB 25 DQAAARTLAQVPTPADLTFYESIGD-VT---WEDVNGCYFLNP----- 65
QY 166 KRLTGEELLTASROOLIMMEADKAVGPIILNSALPDAGWFIADKSGAGERSGITIALG 225
DB 66 ---AGDLLRLQRYGVDFGTDEKSRGLV-----GNSNG-GLIYVAG 104
QY 226 PDGKPRIVVITTTGSOATMDERN-----ROIAEI 255
DB 105 PDG-----AVYKRT-ASLDPAELDKVADGCLQFLEL 135

RESULT 8

hypothetical protein SCJ11.04 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36975

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999

A:Reference number: Z21618

A:Accession: T36975

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-195 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CA52889.1; GSPDB:GN00070; SCOEDB:SCJ11.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.04

Query Match 5.3%; Score 71.5; DB 2; Length 195;
Best Local Similarity 17.6%; Pred. No. 11e+02;
Matches 34; Conservative 35; Mismatches 63; Indels 61; Gaps 7;

QY 56 LSRIDAGQDGLGRRIHYSQNDLVEYSPVTEKHLTDGTVRELCSAATFMSDNTAANLLIT 115
DB 46 VAALGEGEILDRV-----IAREELVEALAVSAETGVTEAGEGEGETALVP 93
QY 116 TIGPELETAFLNMGDHYTRLDREPELNEAIPNDERDTPVAMATTLRLKLLTGEL 173
DB 94 APASAEPEQA-----VHQBGLSVSLSPNNOR-----I 124
QY 174 LTIASRQQLIDMMEADKAVGPIILRSALPA-----GWFIDKSGAGERSGR 218
DB 125 LNVLDORPGLEPYRARDIAALGIEAATAKVEGVREPKRKRLMERGWLQEASGAPFSGRR 184
QY 219 GIITAAIGPPDKPS 231
DB 185 -LVAS--PGDPS 194

RESULT 9

EB1708

conserved hypothetical protein TC0378 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MOpn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: EB1708

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39

A:Reference number: AB1500; MVID:20150255; PMID:10684935

A:Accession: EB1708

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-153 <RET>

A:Cross-references: GB:AE002305; GB:AE002160; NID:g190418; PIDN:AAF9236.1; PID:g719

A:Experimental source: strain Nigg (MOpn)

C:Genetics:

A:Gene: TC0378

Query Match 5.3%; Score 71; DB 2; Length 153;
Best Local Similarity 34.6%; Pred. No. 89;
Matches 18; Conservative 10; Mismatches 16; Indels 8; Gaps 2;

QY 124 TAFILNMGDHYTRLDREPELNEAIPNDERDTPVAMATTLRLKLLTGEL 173
DB 101 TVIABERGHVTLSEFPNTDLNELMODDKRTTP-----RREKLSGFL 146

RESULT 10

AB3753

hypothetical protein BH0825 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: AB3753

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: AB3650; MVID:20512582; PMID:11058132

A:Accession: AB3753

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-177 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04544.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0825
C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match 5.3%; Score 71; DB 2; Length 177;
Best Local Similarity 20.3%; Pred. No. 1.1e+02;
Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

DB 42 PMSTFKVLLCGAVLSRIDAGEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAA 101
11 PEMDTSVGLFYAMVEENV-----RLH-----HLEIDVTEELVYK 47
QY 102 ITMSDNTANLLTTIGPKELTAFLHMGDHYTRLD-RMEPEL-NEAIPND----- 151
DB 48 SDGDSNSMAQLL-----NHLTYVDVRFWRFRKGEALPDSLAEHGP 89
QY 152 -ERPTTPVAMATFLRKLLTGELTLTASROO-----LIDMNEADKVAGPILRSALP 201
DB 90 VDKGCKLPVYTSLSVQELIEKQRYVALLETQCALHDDDLARNPYEENQATIRKGL- 148
QY 202 AGWFIAD 208
DB 149 --MHMAD 153

RESULT 11
683096
conserved hypothetical protein PA4399 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
C:Accession: G83096
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laddig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: G83096
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-192 <STO>
A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AAG07787.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4399
C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
Best Local Similarity 26.0%; Pred. No. 1.2e+02;
Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

DB 53 GAVLSRI-----DAGEQL--GRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSA----- 100
DB 2 GNRLSKIVTRNGDREGTLAGGRVPKSHPRIEAIGAVDELINSQGLLAELEARGAP 61
QY 101 -----AITSMDNTANLLTTIGPKELTAFLHMGDHYTRL-----DRMEPEL-----N 145
DB 62 GLEETVQALAPVQNR-----LFDLGELAMPYRALDETETVARLESCLDRNDELGPKN 116
QY 146 EAIPTDERDTTPVAMATFLRKLLTGELTLTASROQLIDMNEADKVAGPILRSALPACMF 205
DB 117 FILPGSR-----PVAQAHVCSLAR-----SAERKQALDOETTEGVLIRYLNRSLDLF 168
QY 206 IADKSGAGEGSGRTI--AALPD 227
DB 169 VAARAIARRGVAEILWEAANPD 192

RESULT 12
E84217
hypothetical protein Yng0594h [Imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84217

R:Ng, W.V.; Kennedy, S.P.; Mahares, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84217
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-108 <STO>
A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AAG19105.1; GSPDB:GN00138
C:Genetics:
A:Gene: Yng0594h

Query Match 5.2%; Score 70; DB 2; Length 108;
Best Local Similarity 27.5%; Pred. No. 67;
Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

DB 44 MSTFKVLLCGAVLSRIDAGEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAAIT 103
1 MDTVTVELDGLALRLAETD-----RVFVRFALAEVTDVTLRRHGDGV-----GSIT 50
QY 104 MSDNTANLLTTIGPKEL-----TAFHMGDHYTRIDR 139
DB 51 NDDGTDRMARTVPGDSDFAVEPRTSFAVAIVDAATRDR 92

RESULT 13
AB0192
conserved hypothetical protein YP01575 [Imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AB0192
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,
M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0192
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-170 <KOR>
A:Cross-references: GB:AL590842; PIDN:CAC90397.1; PID:g15979615; GSPDB:GN00175
C:Genetics:
A:Gene: YP01575
C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match 5.2%; Score 70; DB 2; Length 170;
Best Local Similarity 21.0%; Pred. No. 1.2e+02;
Matches 22; Conservative 25; Mismatches 44; Indels 14; Gaps 3;

DB 56 LSRIDAGEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAATMSDNTANLLTT 115
DB 70 NKRIDISKGNIG--FNFDNMLI-----KRIYVDKTLQAOGLISGRALTA-----RLVV 115
QY 116 TIGPKELTAFLHMGDHYTRLD-RMEPELNEAIPNDERTTPVA 160
DB 116 DNSGSEFAIITPASYADKIAQRDASSTVLNSALQGEEDGDPYA 160

RESULT 14
D87360
conserved hypothetical protein CC0895 [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87360
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Risen, J.; Heidelberg,
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of Caulobacter crescentus.

A:Accession: A87249; MUID:21173698; PMID:11259647
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-167 <STO>
 A:Cross-references: GB:AE005673; NID:g13422160; PIDN:AAK22880.1; GSFPDB:GN00148
 C:Genetics:
 A:Gene: CC0895

Query Match 5.2%; Score 69.5; DB 2; Length 167;
 Best Local Similarity 25.4%; Pred. No. 1.3e+02;
 Matches 36; Conservative 19; Mismatches 56; Indels 31; Gaps 6;

OY 2 PETLVYKDA-----EDQLGARVGTIELDLN---SGKTIIESFRSEERPPMSTFK 48
 Db 24 PWTLLIRDAFAGVKRFEDQDRLGVARNTAARLKTIVSHGMETRRYSER-PPRHEW 82
 OY 49 VILCGAVLSRIRDAGQOLGRIRIHSQNDLVEYSPYTEKHLTDGMTVR-----ELCSNAIT 103
 Db 83 LFEKGLALSPVLLTMAEMGDHRYGR---DKSPVLFRRHTCGCAPHPVLACEAGQVVD 138
 OY 104 MSDNTANILLTTTIGGPKELTA 125
 Db 139 RRDIERA-----GPDLTA 152

RESULT 15

F84251
 hypothetical protein Vng0953c [Imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84251
 R:Ng, W.V.; Kennedy, S.P.; Mahaitas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Dawson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabcic, Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016930
 A:Accession: F84251
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <STO>
 A:Cross-references: GB:AE004437; NID:g10580510; PIDN:AGI9378.1; GSFPDB:GN00138
 C:Genetics:
 A:Gene: VNG0953C

Query Match 5.1%; Score 68.5; DB 2; Length 145;
 Best Local Similarity 20.3%; Pred. No. 1.3e+02;
 Matches 30; Conservative 28; Mismatches 67; Indels 23; Gaps 3;

OY 41 FPMSTFKVLLCGAVLS-----RIDAGQOLGRIRIHSQNDLVEYSPYTEKH 87
 Db 3 FSVSGSAALFTIAAFVSGILYSAFNGFERVODANMARNDRLTAKNTVEVANTTYDS 62
 OY 88 LFDGTVRELCSAATITMSDNTANILLTTTIGGPKELTAFLHMGDHYTRLDWPEPLNRA 147
 Db 63 VNDITVYATNNGSTTILSYSDVLY---DGETYVDSATVSSVSDGNSQTDLM----- 112
 OY 148 IPNDEBDTTTPVAMATTLKLLTGELLP 175
 Db 113 LPGETYSVYVPTGSAPTRKVVYTGCVT 140

Search completed: September 10, 2003, 12:26:18
 Job time : 11.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_37s_157t
Perfect score: 1345

Sequence: 1 HPETLVKVKDAEDIGARVC.....TMDERNROAIEGASLIKRW 263

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	5.7	191	1 YD4A_SCHPO	Q10346 schizosacch
2	75	5.6	201	1 RACG_DICDI	Q09890 dictyostell
3	68.5	5.1	196	1 RAC2_LOTJA	Q40220 lotus japon
4	68	5.1	122	1 HC3L_THIFE	P80509 thioabactillu
5	68	5.1	149	1 NDK_TREPA	O83974 treponema p
6	66.5	4.9	116	1 RBRP_UREPA	Q09400 ureaplasma p
7	65.5	4.9	206	1 KTHY_METAC	O81899 methanosa
8	65	4.8	177	1 VNSC_RINDR	P35948 rinderpest
9	65	4.8	177	1 VNSC_RINDR	Q03339 rinderpest
10	63.5	4.7	144	1 MARR_ECOLI	P27345 escherichia
11	63.5	4.7	185	1 YCJC_ECOLI	P38522 escherichia
12	63.5	4.7	196	1 ALKH_SACSU	P50846 b kmg/kdpg
13	63	4.7	176	1 HSLV_THEMA	Q09421 thermotoga
14	62.5	4.6	146	1 HBG_RABIT	P02099 oryctolagus
15	62.5	4.6	150	1 PDUV_SALTY	O94006 salmone
16	62.5	4.6	160	1 TATB_HELPY	O25700 helicobacte
17	62.5	4.6	172	1 TATB_HELPY	P31131 escherichia
18	62.5	4.6	182	1 PYRE_STRCO	Q09817 streptomyce
19	62.5	4.6	184	1 HRP_LPEST	P37929 pseudomonas
20	62.5	4.6	195	1 TRPF_THEVO	Q09409 thermoplas
21	62.5	4.6	200	1 RR4_PELNE	O946d9 pellicia nees
22	62	4.6	173	1 Y265_BORBU	O51880 borrelia bu
23	62	4.6	174	1 IRBB_RAT	P17491 ratius norv
24	62	4.6	182	1 RRA_BELCH	O09590 belamanda
25	62	4.6	182	1 RRA_BELCH	O09590 belamanda
26	62	4.6	186	1 RRE_RICPR	Q04608 rickettsia
27	62	4.6	186	1 YCEB_SALTY	P40822 salmone
28	61.5	4.6	178	1 HSLV_RALSO	O943d7 ralsonia s
29	61.5	4.6	193	1 SAR2_LYCSE	P52884 lycopersico
30	61	4.5	121	1 SPCR_HUMAN	P03683 homo sapien
31	61	4.5	184	1 AR12_DROME	O06849 drosophila
32	61	4.5	194	1 RR4_TRIEN	O20234 iris ensata
33	61	4.5	200	1 TATB_CAUCR	Q096t1 caulobacter

34	60.5	4.5	103	1 RS10_NEIGO	P48851 neisseria g
35	60.5	4.5	152	1 YU3_YERPE	O82cf8 yerinia pe
36	60.5	4.5	156	1 BFR_AZOVI	P22759 azotobacter
37	60.5	4.5	160	1 TATB_HELPY	O92m58 helicobacte
38	60.5	4.5	184	1 MLR1_SCHPO	O9un95 schizosacch
39	60	4.5	148	1 CALM_BLAEM	O9hty6 blastoclad
40	60	4.5	161	1 PIN_BPY4	P07068 bacterioph
41	60	4.5	178	1 PYRE_ARCFU	O28533 archaeoglob
42	60	4.5	103	1 UCRI_ANASP	P70758 anabaena sp
43	59.5	4.4	135	1 RS10_NEIMA	O94t21 neisseria m
44	59.5	4.4	135	1 CCRN_PAROL	O57312 paratubercu
45	59.5	4.4	196	1 I196_ASFB7	P27943 african swi

ALIGNMENTS

RESULT 1
YD4A_SCHPO STANDARD: PRT; 191 AA.
AC Q10346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Clf12.04c in chromosome I.
GN SPAC12.04c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rejandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck C., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wandt R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garcon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie M.R., Paulsen J., Potaashkin J.,
RA Shpakowski G.V., Ussery D., Barrell B.G., Nurse P.,
RT The genome sequence of Schizosaccharomyces pombe.
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL; 269944; CAA3808.1; -
DR PIR; T38062; S67447.
DR GenedB_Spombe; SPAC12.04c; -

Query Match	5.7%	Score 77	DB 1	Length 191
Best Local Similarity	22.6%	Pred. No. 26		
Matches	43	Conservative	22	Mismatches 69; Indels 56; Gaps 7
QY	104	MSDNTANILTTTCGGPELTAFAFHNNGDHTRIDRNEPELNEAIPNDERPTTPVMAAT	163	
DB	1	MSVHSLMLMQLNPGSGIDKIALIVN-----VARLD-----PASSKTAQVSMNLN	46	
QY	164	TLRRLTGLLTLTASROOLIDWMEADKACGLRLSALPAGFIADKSGAGERSGRTIAA	223	
DB	47	EFRC-----ILRLGLYKLIYFNFRKDSPEYMSNAINIGCYATE--GLAFAGKQIISI	99	
QY	224	LGPGKRP-----SNIVYTTTGSQATDERRRQI-----	252	
DB	100	-----SKPLEDKLWIMSSRFWLDLTLLTYQLRRTKEDERKHQLDLASNLASLPICHTWS	155	
QY	253	AEGASLIKH 262		
DB	156	VENGAGLKH 165		
RESULT 2				
RACG_DICDI		STANDARD:	PRT:	201 AA.
ID	RACG_DICDI			
AC	OG9PS0:			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	RAS-related protein racg.			
GN	RACG.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.			
OX	NCHI_TaxID=44689;			
RY	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-AA;			
RX	MEDLINE=21127961; PubMed=11222756;			
RA	Riviero F., Diallich H., Glockner G., Noegel A.A.;			
RT	"The Dictyostelium discoideum family of Rho-related proteins.";			
RL	Nucleic Acids Res. 29:1068-1079(2001).			
CC	-1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
CC	EMBL; AF310893; AAC45130.1; -			
DR	HSSP; P21181; 1AM4.			
DR	Dictydb; DD72777; racg.			
DR	InterPro; IPR003578; GTPase_Rho.			
DR	InterPro; IPR001806; Ras_transfmg.			
DR	InterPro; IPR005325; Small_GTP.			
DR	Pfam; PF00071; ras; 1.			
DR	PRINTS; PR00449; RASTRNSFRMNG.			
DR	SMART; SM00174; RHO; 1.			
DR	TIGRFAMs; TIGR0231; small_gtp; 1.			
KW	GTP-binding; Prenylation; Lipoprotein.			
FT	NP_BIND 10 17			
FT	NP_BIND 57 61			
FT	NP_BIND 115 118			
FT	DOMAIN 32 40			
FT	LIPID 198 198			
SO	SEQUENCE 201 AA; 22813 MW; FGE586497CC169 CRC64;			
Query Match	5.6%	Score 75	DB 1	Length 201
Best Local Similarity	22.6%	Pred. No. 41		

```

Matches 36; Conservative 20; Mismatches 59; Indels 44; Gaps 7
OY 44 MSFKEVLLCGAVLSRIDAGEQLGRIRHYGSONDLV-EYSP-YNEKHLLDGMATRELCOSA 101
Db 1 MSTRKVCVVG-----EGGIGKTSMLSTYSNSISNETOPTVGPNST----- 42
OY 102 ITMSDNTCAANILLTTIGGPELFAF-----LHMGDPVTRLDKWEPELN 145
Db 43 LHMNRKRPYNLSMWDTAGOEFEFSRLRSYPQTVFLCLPSLIAPSSFNILDSWGELN 102
OY 146 EALPNDERDITTPVAAATTLKLLTGELL--TLASROOL 182
Db 103 EMCNP-----TPVLWGTQMDLKSNSVNIIDRLCEKROL 135

RESULT 3
RAC2_LOTJA STANDARD; PRT; 196 AA.
AC 040220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAC2-like GTP binding protein RAC2.
GN RAC2.
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
OX NCBI_TaxID=34305;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Gifu / B-129; TISSUE=Root nodules;
RX MEDLINE=97231679; Pubmed=9076991;
RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
RA "Identification of new protein species among 33 different small GTP-
RA binding proteins encoded by cDNAs from Lotus japonicus, and
RA expression of corresponding mRNAs in developing root nodules."
RL Plant J. 11:237-250(1997).
CC -
CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC -
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.lsb-sib.ch/announce).
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: Z73962; CAA98190.1; -.
CC HSSP: P21181; 1AM4.
DR InterPro: IPR003578; GTPase_Rho.
DR InterPro: IPR001806; Ras_Transfmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; ras; 1.
DR PRINTS: PRO0449; RASTRNSFRMG.
DR SMART: SM00174; RHO; 1.
DR TIGRfams: TIGR00231; small_gtp; 1.
DR GTP-binding: Ptenylation; Lipoprotein.
KM NP_BIND 13 20 GTP (BY SIMILARITY).
FT NP_BIND 60 64 GTP (BY SIMILARITY).
FT NP_BIND 118 121 GTP (BY SIMILARITY).
FT DOMAIN 35 43 EFFECTOR REGION (POTENTIAL).
FT LIPID 193 193 GERANYL-GERANYL (POTENTIAL).
SQ SEQUENCE 196 AA: 21704 MW: 88B0A70C4828CD5 CXC66;

Query Match 5.1%; Score 68.5; DB 1; Length 196;
Best Local Similarity 24.1%; Pred. No. 1.3e+02;
Matches 45; Conservative 23; Mismatches 64; Indels 55; Gaps 11;

OY 44 MSTRKVLGAVLSRIDAGEQLGR---HYSONDL-VESPYTERKULTGMYRELCS 99
Db 1 MSTRFKIC-----VTVGDAVGAKTCMLISYTSNTPFTDVP-----TVDFNFS 44

```


QY 100 AATMSDNTAANLLTTIGPKRE-----LTAFLHMGDHYRLDRPE 143
 DB 45 ANVVV-DOSTVNLGIMTAGEEDVNRRLPLSTYRGADVFLAFSLISRASYSKRWIPE 103
 QY 144 LNEAIPNDERDTTPVAMATTLRLTNGELLTL-ASRQOLIDMWEADKVA---GPLRISA 199
 DB 104 LRHAP-----TYPPI-----VLVGTKRLREDRQYLIDHGGAPITTAQGEELKKA 149
 QY 200 LPAGMFI 206
 DB 150 IGAAYVL 156

RESULT 4
 HC3L_THIFE STANDARD; PRT; 122 AA.
 AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
 DE (Fragments).
 GN HOG.
 OS Rhodococcus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 OX NCBI_TaxID=920;
 RN 11
 RP SEQUENCE.
 RC STRAIN-ATCC 19859;
 RX MEDLINE=96241862; PubMed=8661919;
 RA Fischer J., Quentmeyer A., Kostka S., Kraft R., Friedrich C.G.:
 "Purification and characterization of the hydrogenase from
 Rhodococcus ferrooxidans."
 RT Arch. Microbiol. 165:289-296(1996).
 RL -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 - 2 H(+) + 2
 CC ferriocytochrome c3.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 KM Oxioreductase: Iron
 FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 78 79
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EBAFAF365348DA0B CRC64;

Query Match 5.1%; Score 68; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 79;
 Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

QY 19 VGTIELDNLGKILIEFSREERPPMSTFKVLLCGAVLSRIDAGQOLGRRIHYSNDLV 78
 DB 12 VGRVEDDLSXSLIEFRN-----ALIRFGGG-----LG 41
 QY 79 EYSPVTEKHLIDGMVRELCSAATMSDNTAANLLTTIGPKREL 124
 DB 42 KVAFTGTNYEIGVT-----ISGDKNPQAGLVVT-----PREST 75

RESULT 5
 NDK_TREPA STANDARD; PRT; 149 AA.
 ID NDK_TREPA
 AC 083974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 DE (Nucleoside-2-P kinase).
 OS NDK OR TP1010.
 GN Treponema pallidum.
 CC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum R.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.:
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 spirochete."
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate - ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AE001269; AAC65961.1; -
 DR PIR: F71252; F71252.
 DR HSSP: P22887; INPK.
 DR TIGR: TP1010; -
 DR HAMAP: MF_00451; -; 1.
 DR InterPro: IPR001564; NDK.
 DR Pfam: PF00334; NDK; 1.
 DR PRINTS: PR01243; NUCDPKINASE.
 DR PRODOM: PD001018; NDK; 1.
 DR SMART: SM00562; NDK; 1.
 DR PROSITE: PS00469; NDP_KINASES; 1.
 DR KMW Transferase: Kinase; ATP-binding; Complete proteome.
 FT ACT_SITE 117 117
 FT ACT_SITE 117 117
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

Query Match 5.1%; Score 68; DB 1; Length 149;
 Best Local Similarity 21.6%; Pred. No. 1e+02;
 Matches 35; Conservative 25; Mismatches 64; Indels 36; Gaps 7;

QY 29 GILIESFSEERPPMSTFKVLLCGAVLSRIDAGQOLGRRIHYSNDLVETPYTEKHL 88
 DB 21 GEVLSRF---ERKGVLTALRLIC-----VDPAVAEL---HYAEHRRKPPYSLIAYI 67
 QY 89 TDGMTVRELCSAATMSDNTAANLLTTIGPKRELTAFLHNGDHYTRLDRPEPELNAI 148
 DB 68 TSAPVV-----ALAFKGENAISTLVTLGGS-----TVEHAPQ---GTI 103
 QY 149 PND-ERDTPPVAMATTLKLTNGELLTLASRQOLIDMWEAD 189
 DB 104 RGDFALETTNTIVHASSDPSARALALYFSQADVEVERDGN 145

RESULT 6
 RBFA_UREPA STANDARD; PRT; 116 AA.
 ID RBFA_UREPA
 AC 09POH0;

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
OS RBFA OR U0321.
CC Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OK NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3.
RA MEDLINE=20500219; PubMed=11048724;
RA Glas J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caspell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002129; AAF30730.1; -
DR HAMAP: MF_00003; -; 1.
DR InterPro: IPR000338; Rib_bind_factA.
DR Pfam: PF007327; Rib_bind_factA; 1.
DR ProDom: PD007327; Rib_bind_factA; 1.
DR TIGRFAMs: TIGR00082; rbfA; 1.
DR PROSITE: PS01319; RBFA; FALSE_NEG.
DR RNA processing; Complete proteome.
SQ SEQUENCE 116 AA; 13247 MW; A79DC671F0547514 CRC64;
SO

Query Match 4.9%; Score 66.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. No. 98;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

QY 144 LNEALPDEROTTTVAAATTLRLKLTGEL-----LTLASQQLIDWME-ADKVAQPL 196
DB 18 INNALANEINDKIALIARTAVR--LSNDLSVAKIFLDAHKRESMLKYLENVNKKVSG-L 74

QY 197 RSALPAGW-----FIADKS 210
DB 75 RSKLAEMTSYKVPFLRFVIDET 97

RESULT 7
KTHY_METAC
ID KTHY_METAC STANDARD; PRT; 206 AA.
AC 08THS9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
OS TMK OR MA4433.
CC Methanosarcina acetivorans.
CC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
CC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OK NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endlitz M.G., MacDonald P.,

```

```

RA FitzHugh W., Calvo S., Engels R., Smitrov S., Alnoor D., Brown A.,
RA Allen N., Meylor J., Strange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McKeen P., McKernan K., Talmas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnet H.C., Kirycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umeyan L.A., White O., White R.H., de Maccario E.C.,
RA Ferry J.G., Jarrell K.E., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.,
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE011164; AAM07774.1; -
DR HAMAP: MF_00165; -; 1.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR TIGRFAMs: TIGR00041; dTMP_Kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
DR Translasee; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C25B68542383 CRC64;
SO

Query Match 4.9%; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.7%; Pred. No. 2,46+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

QY 56 LSRID-AGOEOLGRINHSNDLVEYSPTKHLTDCMTVELCSAAITMSDNTAAILL 114
DB 8 LEGIDGSKSVYAKKIO-ENSELAVFEVFIREPTRG-----TLGNNAVENAIO 55

QY 115 TTIGCPHELTAFELNMGDVTYRLDRWPELNEAIPNDEROTTTVAAATTLRLKLTGEL 174
DB 56 SDTQIAELFLFTADHEHLAKLVKPALEDKTIVSDRYSRYAOGITITLKNRLDNP- 114

QY 175 TLASRQQLIDWMEADKVAQPLRSALPAGW-----FIAD-----KSG-AGE--- 214
DB 115 -----EWR-----DLHGRWTVIPDLTFLFDIEPIAVKRCGRGBOTKF 154

QY 215 -----RSGRCIIAALGPDGKPSRIWVYTTGSOATNDE 247
DB 155 EKIEFLRGVRELFLGLAAE-EPEHFVIVDASGSPEDYER 192

RESULT 8
VNSC_RINDK
ID VNSC_RINDK STANDARD; PRT; 177 AA.
AC P35948;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein C.
OS C.
CC Rinderpest virus (strain Kabete O) (RDV).
CC Viruses; ssRNA negative strand viruses; Mononegavirales;
CC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OK NCBI_TaxID=11242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92410646; PubMed=1529555;

```


RT "Characterization of MarR, the repressor of the multiple antibiotic
 resistance (mar) operon in *Escherichia coli*.";
 RL J. Bacteriol. 177:3414-3419(1995).
 RN [6]
 RP MUTAGENESIS.
 RX MEDLINE-20223625; PubMed-10760140;
 RA Aleksun M.N., Kim Y.S., Levy S.B.;
 RT "Mutational analysis of MarR, the negative regulator of marRAB
 expression in *Escherichia coli*, suggests the presence of two regions
 required for DNA binding.";
 RL Mol. Microbiol. 35:1394-1404(2000).
 CC -1- FUNCTION: REPRESSOR OF THE MARAB OPERON WHICH IS INVOLVED IN THE
 CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS
 CC GENES. BINDS TO THE MARO OPERATOR/PROMOTER SITE.
 CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M96235; AAC16394.1; ALT_INIT.
 DR EMBL; AE000250; AAC74603.1; ALT_INIT.
 DR EMBL; D90795; BAA15212.1; -
 DR EMBL; D90796; BAA15220.1; -
 DR EMBL; D90797; BAA15232.1; -
 DR PIR; E64907; E64907.
 DR PDB; 1JGS; 28-DEC-01.
 DR Ecogene; EG11435; marR.
 DR InterPro: IPR000835; HTH_MarR.
 DR Pfam: PF01047; MarR_1.
 DR PRINTS; PRO0598; HTHMAR.
 DR SMART; SM00347; HTHMAR.
 DR PROSITE; PS01117; HTH_MAR_FAM1LY; 1.
 DR Transcription regulation; DNA-binding; Repressor;
 DR Antibiotic resistance; Complete proteome; 3D-structure.
 FT MUTAGEN 45 45
 FT REGION II TRANSCRIPT.
 FT MUTAGEN 77 77
 FT R->L: INCREASED TRANSCRIPTION OF THE
 FT REGION II TRANSCRIPT.
 FT MUTAGEN 123 144
 FT MISSING: INCREASED TRANSCRIPTION OF THE
 FT REGION II TRANSCRIPT.
 SQ SEQUENCE 144 AA; 16065 MW; BE7DF5549E24D1D3 CNG64;
 Query Match 4.7%; Score 63.5; DB 1; Length 144;
 Best Local Similarity 29.5%; Pred. No. 2.2e+02;
 Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;
 OY 66 IGRRIHY--SOND--LVEY-SP--VTEKHLTDGMVRELCSAIVMSNTANLLLTIG 118
 DB 14 IGRLLHMYNOKRDLNELYSLFDITMAOFVLCIR--CAACIT----- 56
 OY 119 GKRLTAFRLH-NMGDVTFLDR-----WEPELNEAIPDNDER-----TTTPVAAATTLRK 167
 DB 57 -PVELKTVSLVDGLATRLDLVCKGWERLPN--PDKRGVLTGKTGGALICEQCHQ 113
 OY 168 LITGELTLTASNOOLIDMHEADKVA--GPLRSALP 201
 DB 114 LVGGDL-----HOELTKNLTADEVATLEYLELTKKVP 144
 RESULT 11
 YCJC_ECOLI
 ID YCJC_ECOLI STANDARD; PRT; 185 AA.
 AC P38522; P76839; P77417;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein ycJc.

GN YCJC OR B1299 OR SF1304.
 OS *Escherichia coli*, and
 OS *Shigella flexneri*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-E.coli; STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-E.coli; STRAIN-K12;
 RX MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Bada T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio T., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takada J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 65-185 FROM N.A.
 RC SPECIES-E.coli;
 RX MEDLINE-91216440; PubMed-1840553;
 RA Helm R., Strehler E.E.;
 RT "Cloning an *Escherichia coli* gene encoding a protein remarkably
 RT similar to mammalian aldehyde dehydrogenases.";
 RL Gene 99:15-23(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S.flexneri; STRAIN-301 / Serotype 2a;
 RX MEDLINE-22272406; PubMed-12384590;
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity
 RT through comparison with genomes of *Escherichia coli* K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [5]
 RP IDENTIFICATION.
 RC SPECIES-E.coli;
 RA Rudd K.E.;
 RT Unpublished observations (Aug-1994).
 RL -1- SIMILARITY: SOME, TO H.INFLUENZAE H10559.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AE000228; AAC74381.1; -
 DR EMBL; D90768; BAA14868.1; -
 DR EMBL; D90767; BAA14859.1; -
 DR EMBL; M38433; -, NOT_ANNOTATED_CDS.
 DR EMBL; AE015157; AAN42915.1; -
 DR PIR; F64878; F64878.
 DR Ecogene; EG12431; ycJc.
 DR InterPro: IPR007113; Cupin_sup.
 DR InterPro: IPR001387; HTH_3.

DR Pfam: PF01381; HTH_3; 1.
 RA SMART: SM00530; HTH_XRE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 185 AA; 20091 MW; D7DD03FD794A7768 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 63.5; DB 1; Length 185;
 Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;

OY 145 NEAIPNDEPDTTPVAMATTLRLRLTGLLETLASROOLIDWMEADRVAGPLRSALPAGW 204
 DB HSAISTITIEDKVSRA--ISTIAKRLKVGSLSE-----FESPEKPERDEQV----- 78
 OY 205 FIADKSGAGRGSRGIIAALGPGKRSRIYV-----YTTGSOATYMDR-NRQIAETGASL 259
 DB 79 -VINODLLIEMSGSGVSMKLVHNGNPRLAMIFETYPGT--TGERIKHQEEIGTVL 135

RESULT 12
 ALKH_BACSU STANDARD; PRT; 196 AA.
 AC P50846;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE KHG/KDPG aldolase [Includes: 4-hydroxy-2-oxoglutarate aldolase
 (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase (KHG-aldolase); 2-
 dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
 de aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
 aldolase)].
 GN KDG.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1423;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azavedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 Serror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 the serA and kds loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 Azavedo V., Berrero M.G., Bessieres P., Bolotin A., Borchert S.,
 Bortles R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 Denton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,
 Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogilawa A., Oudega B., Park S.H.,
 Pirescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 Rieger M., Rivolet C., Rocha E., Roche B., Rose M., Sadate Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
 Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vasarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yeta K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Darchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis".
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate - pyruvate +
 glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate -
 pyruvate + D-glycerolaldehyde 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTER-DODUOPOFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 OF GLYOXYLATE.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: LA7838; AAB38480.1; -
 DR EMBL: 299115; CAB14127.1; -
 DR PIR: H69647; H69647.
 DR Subtilist; BG11396; Kdga.
 DR InterPro: IPR000887; Aldlase_KDPG_KHG.
 DR Pfam: PF01081; Aldolase; 1.
 DR TIGRFAMs: TIGR01182; eda; 1.
 DR PROSITE: PS00159; ALDOLASE_KDPG_KHG_1; FALSE_NEG.
 DR PROSITE: PS00160; ALDOLASE_KDPG_KHG_2; 1.
 KW Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
 FT ACT_SITE 43 43 By similarity.
 FT ACT_SITE 47 47 By similarity.
 FT BINDING 130 130 SCHIFF-BASE WITH KHG OR PYRUVATE (BY
 FT SIMILARITY).
 SQ SEQUENCE 196 AA; 20865 MW; EBOB597FD1102C4 CRC64;

Query Match
 Best Local Similarity 4.7%; Score 63.5; DB 1; Length 196;
 Matches 27; Conservative 17; Mismatches 59; Indels 19; Gaps 4;

OY 31 IIESPSEPFPMSTFKYLGAVALSRIDAGO-EQGRIRYSON---DLYEYSPYER 86
 DB 54 IIESRNREDI-----LIGAGYVISAQOAGEAMAGQFIVSPFSADIAEHLSPYRT 106
 OY 87 HLTDMTVRELCSAATMSDNTAANL-----LTTTGCPRELTAFLHNGDHYTRLD 138
 DB 107 HVIPGLVRLSEIMEALTFEFTTLKLPSSGVFGIPIPKMKNLAGFPQVTFIPGCIHPSEVP 166
 OY 139 RW 140
 DB 167 DW 168

RESULT 13
 HSLV_THEMA STANDARD; PRT; 176 AA.
 AC Q9WY21;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 28-FEB-2003 (Rel. 41, last annotation update)
 DE ATP-dependent protease hslV (EC 3.4.25.-).
 GN HSLV OR TM0521.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxId=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;


```

BC STRAIN-LTf2 / SGCs1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.†;
RL Nature 413:852-856(2001).
CC -1- INDUCTION: By propenadiol.
CC -1- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF026270; AAD39020.1; -.
DR EMBL; AE008790; AAL20960.1; -.
DR dStyGene; SGI0680; pduv.
RV ATP-binding; Complete proteome.
PT NP_BIND 8 15 ATP (POTENTIAL).
SQ SEQUENCE 150 AA; 16348 MW; 77711229432F97E56 CRC64;

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_37S_157T

Perfect score: 1345

Sequence: 1 HPEITLVKVADEQLEARVG.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	534	39.0	105	2	052026	052026 pseudomonas
2	534	39.0	105	2	052330	052330 escherichia
3	499	37.1	145	2	08RTD8	08RTD8 klebsiella
4	488	36.3	102	2	052639	052639 pseudomonas
5	464	34.5	138	2	0990F3	0990F3 escherichia
6	464	34.5	138	2	09AMA1	09AMA1 escherichia
7	464	34.5	138	2	09AMA2	09AMA2 escherichia
8	464	34.5	138	2	09AMA0	09AMA0 escherichia
9	462	34.3	138	2	09AM99	09AM99 escherichia
10	462	34.3	139	2	09AM98	09AM98 escherichia
11	459	34.1	139	2	09AM93	09AM93 escherichia
12	338	25.1	67	2	053553	053553 shigella fl
13	326	24.2	95	2	08GDE5	08GDE5 klebsiella
14	320	23.0	62	2	08JMS8	08JMS8 shigella fl
15	301	22.4	128	2	08VQ00	08VQ00 escherichia
16	251	18.7	180	2	08KVT2	08KVT2 staphylococ

17	242	18.0	52	2	09R412	09R412 shigella fl
18	241	17.9	48	2	09RLH0	09RLH0 proteus mir
19	197	14.6	38	2	P97145	P97145 escherichia
20	173.5	12.9	134	2	08VUD3	08VUD3 staphylococ
21	128.5	9.6	109	2	053698	053698 staphylococ
22	105	7.8	100	2	P97146	P97146 escherichia
23	87	6.5	100	2	093505	093505 staphylococ
24	82	6.1	68	2	09XB2	09XB2 bacillus ce
25	80.5	6.0	202	16	09AC820	09AC820 rhizobium l
26	80	5.9	198	2	09ACM8	09ACM8 streptococ
27	78	5.8	153	5	08S20	08S20 clostridium
28	75	5.6	181	16	08D108	08D108 yersinia intes
29	74.5	5.5	113	2	049970	049970 mycobacteri
30	74.5	5.5	172	16	08BWM3	08BWM3 brucella me
31	74.5	5.5	172	16	08BWM3	08BWM3 brucella me
32	73	5.4	131	2	09X9H0	09X9H0 yersinia en
33	73	5.4	175	17	08BTH8	08BTH8 pyrobaculum
34	73	5.4	201	17	08TW84	08TW84 methanosarc
35	72.5	5.4	145	2	005984	005984 staphylococ
36	72.5	5.4	184	5	045379	045379 caenorhabdl
37	72.5	5.4	200	16	097JEB	097JEB clostridium
38	72	5.4	205	6	09N275	09N275 ovis aries
39	72	5.4	205	16	09B0T8	09B0T8 rhizobium l
40	71.5	5.3	152	16	09R191	09R191 streptomyce
41	71.5	5.3	194	2	08GDC0	08GDC0 borrelia af
42	71.5	5.3	195	16	09RTA0	09RTA0 streptomyce
43	71.5	5.3	196	2	09AH34	09AH34 pseudomonas
44	71	5.3	150	10	09ATG4	09ATG4 nicotiana t
45	71	5.3	153	16	09PRT4	09PRT4 chlamydia m

ALIGNMENTS

RESULT 1

ID	052026	PRELIMINARY;	PRT;	105 AA.
AC	052026;			
DT	01-NOV-1996 (TREMblrel. 01, Created)			
DT	01-NOV-1996 (TREMblrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMblrel. 20, Last annotation update)			
DE	Beta-lactamase (Fragment).			
GN	Bla.			
OS	Pseudomonas aeruginosa.			
OC	Plasmid PR01614.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC	Pseudomonadaceae; Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95011664; PubMed=7926843;			
RA	West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;			
RT	*Construction of improved Escherichia-Pseudomonas shuttle vectors			
RT	derived from pUC18/19 and sequence of the region required for their			
RT	replication in Pseudomonas aeruginosa.*;			
RL	Gene 148:81-86(1994).			
DR	EMBL; L30112; AAA66058.1; -			
DR	HSSP; P00810; IXPB.			
DR	InterPro; IPR001466; Beta_lactamase.			
DR	InterPro; IPR000871; Beta_lactamase_A.			
DR	Pfam; PF00144; beta-lactamase; 1.			
DR	PRINTS; PR00118; BLACTAMASEA.			
KW	Plasmid.			
FT	NON_TER			
SO	SEQUENCE	1	1	
		105 AA;	11229 MW;	D28899A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 2.2e-37;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMATRLKLTGELTLASROOLDMMEDAKVAGPILRSALPAGWRTADKSGAGGSGSG 219

DB 2 AAMATRLKLTGELTLASROOLDMMEDAKVAGPILRSALPAGWRTADKSGAGGSGSG 61

QY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 105

RESULT 2

Q52330 PRELIMINARY; PRT; 105 AA.

ID Q52330
 AC Q52330:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Tnl bla protein (Fragment).
 OS Escherichia coli.
 OC Plasmid RK2.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-90264294; PubMed-2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The *kil*-kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT *korc*.";
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL: M32794; AAA26408.1; -;
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase.A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 2.2e-37;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 219
 DB 2 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 61

QY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 105

RESULT 3

Q8RTD8 PRELIMINARY; PRT; 145 AA.

ID Q8RTD8
 AC Q8RTD8:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Klebsiella.
 CX NCBI_TaxID=573;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Palaeobacterium S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftazidime-resistant Klebsiella pneumoniae.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF467105; AAL75506.1; -;
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER
 SQ SEQUENCE 145 AA; 145 MW; DA95AF7557DA13D4 CRC64;

QY 160 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 219
 DB 2 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 61

QY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGASLTKHW 105

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 37.1%; Score 499; DB 2; Length 145;
 Best Local Similarity 68.8%; Pred. No. 4.8e-35;
 Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 100 AAITMSDNTAANLLTTTIGSPKELTAFLNHGDHVTIRIDRNEPELNEAIPNDESDTTPV 159
 DB 7 AAXYMSDMSAANLLATLVGPGGLTAFLRQIGDVTIRIDRNETELNEALPGDARDTTPA 66

QY 160 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 219
 DB 67 SMAATLRKLTLTSQRLSARSQKQLQWYDNDVAGPLIRSVLPAGFIADKSGAGERSG 126

QY 220 IIAALGPDGKPSRIYVIT 237
 DB 127 IVALGPNKNAERIVIT 144

RESULT 4

Q52639 PRELIMINARY; PRT; 102 AA.

ID Q52639
 AC Q52639:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Beta-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 CC Plasmid pRO1600.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 CX NCBI_TaxID=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn1.
 RX MEDLINE-94336757; PubMed-8058819;
 RA Jansons I., Touchle G., Sharp R., Almqvist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the *pOF*
 RT *pRO1600* *oriR* region found in the broad-host-range plasmids of the *pOF*
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL: U22691; AAA98312.1; -;
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase.A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 2.5e-34;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 219
 DB 2 AMATTLRKLTLGELLTLASROQLIDWMEADKVAGPLLSALPAGFIADKSGAGERSG 61

QY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGA 257
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNQIAEIGA 99

RESULT 5

Q99QF3 PRELIMINARY; PRT; 138 AA.

ID Q99QF3
 AC Q99QF3:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6988, and E/99 5-2;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant Escherichia coli
 RT isolates from UMMC, Malaysia.
 RL EMBL: AF327049; AAK07464.1; -
 DR EMBL: AF327051; AAK07466.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 138;
 Best Local Similarity 68.5%; Pred. No. 4.5e-32;
 Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKKELTAFLNMGDHYTRLDREPELNEAIPNDERDTTPVAMATLTKLLT 170
 DB 10 NLLATVGGPAGLTAFLEQIDNVTRLDRETELNEALPDARDTTTPASMAATLTKLLT 69
 OY 171 GELLTASRQQLIDMNEADKVAPLRSALPAGMFIADSGAGERSGRTIAALGPDKP 230
 DB 70 SORLSARSORQLQMWVDVAGPLIRSVLPAGMFIADKTGASKRGARGIVALGPNKKA 129
 OY 231 SRIVIVY 237
 DB 130 ERIVIVY 136

RESULT 6

O9AMA1 PRELIMINARY; PRT; 138 AA.
 ID O9AMA1;
 AC O9AMA1;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=18/25;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant Escherichia coli
 RT isolates from UMMC, Malaysia.
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327050; AAK07465.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;
 OY 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;
 Query Match 34.5%; Score 464; DB 2; Length 138;
 Best Local Similarity 68.5%; Pred. No. 4.5e-32;
 Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKKELTAFLNMGDHYTRLDREPELNEAIPNDERDTTPVAMATLTKLLT 170
 DB 10 NLLATVGGPAGLTAFLEQIDNVTRLDRETELNEALPDARDTTTPASMAATLTKLLT 69
 OY 171 GELLTASRQQLIDMNEADKVAPLRSALPAGMFIADSGAGERSGRTIAALGPDKP 230
 DB 70 SORLSARSORQLQMWVDVAGPLIRSVLPAGMFIADKTGASKRGARGIVALGPNKKA 129
 OY 231 SRIVIVY 237
 DB 130 ERIVIVY 136

RESULT 7

O9AMA2 PRELIMINARY; PRT; 139 AA.
 ID O9AMA2;
 AC O9AMA2;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=918;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant Escherichia coli
 RT isolates from UMMC, Malaysia.
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327048; AAK07463.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;
 OY 139 AA; 15234 MW; 0B161D09282442A8 CRC64;
 Query Match 34.5%; Score 464; DB 2; Length 139;
 Best Local Similarity 68.5%; Pred. No. 4.6e-32;
 Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKKELTAFLNMGDHYTRLDREPELNEAIPNDERDTTPVAMATLTKLLT 170
 DB 10 NLLATVGGPAGLTAFLEQIDNVTRLDRETELNEALPDARDTTTPASMAATLTKLLT 69
 OY 171 GELLTASRQQLIDMNEADKVAPLRSALPAGMFIADSGAGERSGRTIAALGPDKP 230
 DB 70 SORLSARSORQLQMWVDVAGPLIRSVLPAGMFIADKTGASKRGARGIVALGPNKKA 129
 OY 231 SRIVIVY 237
 DB 130 ERIVIVY 136

RESULT 8

O9AMA0 PRELIMINARY; PRT; 138 AA.
 ID O9AMA0;
 AC O9AMA0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 OY 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;
 Query Match 34.5%; Score 464; DB 2; Length 138;
 Best Local Similarity 68.5%; Pred. No. 4.5e-32;
 Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

RC STRAIN-E/98 9-1;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant *Escherichia coli*
 RL Isolates from UMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327052; AAK07467.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; Beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1 1
 FT 138 138
 SO SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 6.7e-32;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLLHNGDHYTRLDREPELNEAIPNDERDITTPVAMATTLRLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDADDTTPASMAATLRLKLLT 69
 OY 171 GELLTLASROQLIDMNEADKVAGPLLRSGALPAGMFIADKSGAGERSGIIAAGPDKRP 230
 DB 70 SORLSARSORQLQMWVDVAVAGPLIRSVLPAGMFIADKTGASKRGARIVALLGPNKKA 129
 OY 231 SRIIVYI 237
 DB 130 ERIIVLY 136

RESULT 9

ID O9AM99 PRELIMINARY; PRT; 138 AA.
 AC O9AM99;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Escherichia*.
 CX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/99 4-1;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant *Escherichia coli*
 RL Isolates from UMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327053; AAK07468.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; Beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1 1
 FT 138 138
 SO SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 6.7e-32;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLLHNGDHYTRLDREPELNEAIPNDERDITTPVAMATTLRLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDADDTTPASMAATLRLKLLT 69
 OY 171 GELLTLASROQLIDMNEADKVAGPLLRSGALPAGMFIADKSGAGERSGIIAAGPDKRP 230
 DB 70 SORLSARSORQLQMWVDVAVAGPLIRSVLPAGMFIADKTGASKRGARIVALLGPNKKA 129

OY 231 SRIIVYI 237
 DB 130 ERIIVLY 136

RESULT 10

ID O9AM98 PRELIMINARY; PRT; 139 AA.
 AC O9AM98;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Escherichia*.
 CX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/99 3-2;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant *Escherichia coli*
 RL Isolates from UMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327054; AAK07469.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; Beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1 1
 FT 139 139
 SO SEQUENCE 139 AA; 15234 MW; 03361A792F5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
 Best Local Similarity 67.7%; Pred. No. 6.8e-32;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLLHNGDHYTRLDREPELNEAIPNDERDITTPVAMATTLRLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDADDTTPASMAATLRLKLLT 69
 OY 171 GELLTLASROQLIDMNEADKVAGPLLRSGALPAGMFIADKSGAGERSGIIAAGPDKRP 230
 DB 70 SORLSARSORQLQMWVDVAVAGPLIRSVLPAGMFIADKTGASKRGARIVALLGPNKKA 129
 OY 231 SRIIVYI 237
 DB 130 ERIIVLY 136

RESULT 11

ID O9AM93 PRELIMINARY; PRT; 139 AA.
 AC O9AM93;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; *Escherichia*.
 CX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-935;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from ceftazidime-resistant *Escherichia coli*
 RL Isolates from UMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327047; AAK07462.1; -
 DR HSSP: P14557; 1SHV.
 DR InterPro: IPR001466; Beta_lactamase.

OY 111 NLLTTTGGPKELTAFLLHNGDHYTRLDREPELNEAIPNDERDITTPVAMATTLRLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDADDTTPASMAATLRLKLLT 69
 OY 171 GELLTLASROQLIDMNEADKVAGPLLRSGALPAGMFIADKSGAGERSGIIAAGPDKRP 230
 DB 70 SORLSARSORQLQMWVDVAVAGPLIRSVLPAGMFIADKTGASKRGARIVALLGPNKKA 129

[illegible]

CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK51942 to AAK54550 and AAM82169
CC represent sequences used in the exemplification of the present invention.

39

XX PS Claim 20: SEQ ID No 58276; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantifying a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp://wipo.int/pub/published_pcr_sequences](http://wipo.int/pub/published_pcr_sequences).

[illegible]

PR 18-AUG-2000: 2000US-0226279.
 PR 22-AUG-2000: 2000US-0226581.
 PR 22-AUG-2000: 2000US-0226688.
 PR 22-AUG-2000: 2000US-0227182.
 PR 23-AUG-2000: 2000US-0227009.
 PR 30-AUG-2000: 2000US-0228924.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229343.
 PR 01-SEP-2000: 2000US-0229344.
 PR 01-SEP-2000: 2000US-0229345.
 PR 05-SEP-2000: 2000US-0229509.
 PR 06-SEP-2000: 2000US-0230437.
 PR 06-SEP-2000: 2000US-0230438.
 PR 08-SEP-2000: 2000US-0231242.
 PR 08-SEP-2000: 2000US-0231243.
 PR 08-SEP-2000: 2000US-0231244.
 PR 08-SEP-2000: 2000US-0231413.
 PR 08-SEP-2000: 2000US-0231414.
 PR 08-SEP-2000: 2000US-0232080.
 PR 12-SEP-2000: 2000US-0232081.
 PR 14-SEP-2000: 2000US-0231968.
 PR 14-SEP-2000: 2000US-0232397.
 PR 14-SEP-2000: 2000US-0232398.
 PR 14-SEP-2000: 2000US-0232399.
 PR 14-SEP-2000: 2000US-0232400.
 PR 14-SEP-2000: 2000US-0233401.
 PR 14-SEP-2000: 2000US-0233063.
 PR 14-SEP-2000: 2000US-0233064.
 PR 21-SEP-2000: 2000US-0233065.
 PR 21-SEP-2000: 2000US-0234274.
 PR 25-SEP-2000: 2000US-0234997.
 PR 25-SEP-2000: 2000US-0234998.
 PR 26-SEP-2000: 2000US-0235484.
 PR 27-SEP-2000: 2000US-0235484.
 PR 27-SEP-2000: 2000US-0235834.
 PR 29-SEP-2000: 2000US-0235836.
 PR 29-SEP-2000: 2000US-0236327.
 PR 29-SEP-2000: 2000US-0236367.
 PR 29-SEP-2000: 2000US-0236368.
 PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239935.
 PR 13-OCT-2000: 2000US-0239937.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241221.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.
 PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.
 PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.

PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249212.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249246.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250160.
 PR 03-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 06-DEC-2000: 2000US-0256719.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251899.
 PR 11-DEC-2000: 2000US-0251990.
 PR 05-JAN-2001: 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Rosen CA, Barnash SC, Ruben SM;
 DR WPI: 2001-465566/50.
 XX N-PSDB: AAS41090.
 DR
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 PT
 XX
 PS
 XX
 CC
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. hemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU29915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 129 AA:

Query Match 34.9%; Score 469; DB 22; Length 129;
 Best Local Similarity 95.8%; Pred. No. 1,4e-40;
 Matches 92; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

SQ	Sequence	146 AA:
Query Match	41.9%;	Score 563; DB 22; Length 146;
Best Local Similarity	96.5%;	Pred. No. 2.8e-50;
Matches 111;	Conservative 1;	Mismatches 3; Indels 0; Gaps 0
Oy		
Dd		
Oy		
Dd		
RESULT 4 AAW20440 ID AAW20440 standard; protein; 127 AA. XX AC AAW20440; XX DT 14-JUL-1997 (first entry) DX H. pylori cytoplasmic protein, 34574062_aa. XX Cytoplastic; vaccine; prevention; treatment; infection; identification; KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor; KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome; KW replication; transcription; recombination; repair. <br/ XX Helicobacter pylori. OS Key Location/Qualifiers FH MISC-difference 121 /note= "encoded by codon YGC" FT Misc-difference 122 /note= "encoded by codon WGG" ET WO9640893-A1. PN 19-DEC-1996. PP 06-JUN-1996; 96MO-USO9122. PR 01-APR-1996; 96US-0630405. PX 07-JUN-1995; 95US-0487032. PA (ASTRA) ASTRA AB. PI Berglindh OT, Smith D, Mellgaard BL; XX MPI; 1997-052306/05. DR N-PADB; AAT67613. XX The present sequence is a Helicobacter pylori cytoplasmic protein CC involved in genomic replication, transcription, recombination and repair CC The protein may be used in a vaccine to prevent or treat H. pylori CC Infection or to identify H. pylori polypeptide binding compounds, useful CC as potential H. pylori life cycle activators or inhibitors. The genomic CC sequence of H. pylori (ATCC 55679) was determined from overlapping CC contigs generated by mechanically shearing the bacterial DNA. The CC sequences were analysed for ORF of at least 180 nucleotides, and the CC predicted coding regions defined by computer evaluation. To identify CC likely H. pylori antigens for vaccine development, the amino acid CC sequences predicted from various ORFs were analysed for significant CC homology to other known or exported membrane proteins. Having identified CC and determined the sequences of interest, particular regions can be		

CC	Isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.					
CC						
XX	SQ	Sequence	127 AA;			
OY	Query Match	36.0%; Score 484; DB 18; Length 127;				
DB	Best Local Similarity	97.9%; Pred. No. 3,7e-42;				
XX	Matches	95; Conservative 1; Mismatches 1; Indels 0; Gaps				
OY	1 HPETLVKVKAEADQLGARVGIIELDLSNGEILSFNSSEERFPMSTFKYLICGAVLSRID	60				
DB						
XX	24 HPETLVKKKAEDDLGARVGIIEIDLSNGSKLISFPREERFPMSTFKYLICGAVLSRID	83				
OY	61 AGQEOLGRRIHYSONDLVEYSPTVEKHLLTGMTVREL	97				
DB						
XX	84 AGQEOLGRRIHYSONDLVEYSPTVEKHLLTGMTVREL	120				
RESULT 5						
AAU23220						
ID	AAU23220 standard; Protein: 129 AA.					
XX						
AC	AAU23220;					
XX						
DT						
XX						
DE	18-DEC-2001 (first entry)					
XX						
KM	Novel human enzyme polypeptide #306.					
KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;					
KW	ligase; hyperproliferative disorder; immunodeficiency disorder;					
KW	autoimmune disorder; neurological disorder; metabolic disorder;					
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;					
KM	blood-related disorder; infectious disorder; cytostatic; anti arthritic;					
XX	nephrotropic; anticoagulant.					
OS	Homo sapiens.					
XX						
PN	WO200153301-A2.					
XX						
PD	02-AUG-2001.					
Pf	17-JAN-2001; 2001MO-USO1239.					
XX						
PR	31-JAN-2000; 2000US-0179065.					
PR	04-FEB-2000; 2000US-0180628.					
PR	24-FEB-2000; 2000US-0184664.					
PR	02-MAR-2000; 2000US-0186350.					
PR	16-MAR-2000; 2000US-0189874.					
PR	17-MAR-2000; 2000US-0190076.					
PR	18-APR-2000; 2000US-0198123.					
PR	19-MAY-2000; 2000US-0205515.					
PR	07-JUN-2000; 2000US-0209467.					
PR	28-JUN-2000; 2000US-0214886.					
PR	30-JUN-2000; 2000US-0215135.					
PR	07-JUL-2000; 2000US-0216647.					
PR	07-JUL-2000; 2000US-0216880.					
PR	11-JUL-2000; 2000US-0217487.					
PR	11-JUL-2000; 2000US-0217496.					
PR	11-JUL-2000; 2000US-0218290.					
PR	26-JUL-2000; 2000US-0220963.					
PR	26-JUL-2000; 2000US-0220964.					
PR	14-AUG-2000; 2000US-0224518.					
PR	14-AUG-2000; 2000US-0224519.					
PR	14-AUG-2000; 2000US-0225213.					
PR	14-AUG-2000; 2000US-0225214.					
PR	14-AUG-2000; 2000US-0225266.					
PR	14-AUG-2000; 2000US-0225267.					
PR	14-AUG-2000; 2000US-0225268.					
PR	14-AUG-2000; 2000US-0225270.					
PR	14-AUG-2000; 2000US-0225447.					
PR	14-AUG-2000; 2000US-0225757.					
PR	14-AUG-2000; 2000US-0225758.					
PR	14-AUG-2000; 2000US-0225759.					

[illegible]

PA (EART) EARTH CHEM CO LTD.
XX
PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX
DR WPI: 1986-015031/03.
XX N-PSDB: AAN60631.
XX
PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
and transformed cells contg. it.
XX
XX
PS Disclosure: Page 56-59; 92pp: German.
XX
XX The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small ants. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
CC
SQ Sequence 196 AA:

Query Match 43.0%; Score 579; DB 7; Length 196;
Best Local Similarity 97.4%; Pred. No. 9.3e-52;
Matches 114; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPEITLVKYNDEQIGARVGYIELDLNSGELIESFRSEFRPPMSTFYVLCGAVLSRID 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
24 HPEITLVKYNDEQIGARVGYIELDLNSGELIESFRSEFRPPMSTFYVLCGAVLSRID 83
QY 61 AGQEQIGRRIRHSQNDLVEYSPYRKHTDGTGTVRELCSAAITMSDNTAAILLTTI 117
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
84 AGQEQIGRRIRHSQNDLVEYSPYRKHTDGTGTVRELCSAAITMSDNTAAILLTTI 140

RESULT 2
AAU23221
ID AAU23221 standard; Protein, 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WC200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235485.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0235802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345
Sequence: 1 HPELVKVKAEADQLGARVC.....TMDRRNQIAETGASLIRHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579	43.0	196	7 AAP60627	Beta-urogastrone -
2	563	41.9	146	22 AAU33221	NOVEL human enzyme
3	563	41.9	146	22 AAW90870	Human immune/haema
4	484	36.0	127	18 AAW20440	H. pylori cytoplasm
5	469	34.9	129	22 AAU33220	NOVEL human enzyme
6	469	34.9	129	22 AAW90871	Human immune/haema
7	469	34.9	182	22 ABG27917	NOVEL human diageno
8	457	34.0	94	21 AAW90502	Breast and ovarian
9	396	29.4	159	7 AAP60628	Beta-urogastrone -

10	366	27.2	101	22 ABG27935	NOVEL human diageno
11	325	24.2	88	22 ABG27919	NOVEL human diageno
12	191	14.2	119	7 AAP60626	Beta-urogastrone -
13	141	10.5	51	21 AAY92783	Vtgs-beta-lactama
14	80.5	6.0	158	22 AAW5138	Propionibacterium
15	79	5.9	200	18 AAW55524	H. pylori ORF 29ep
16	75	5.6	202	24 AAE32763	Human zcyto24 prot
17	75	5.6	202	24 AAO16283	Mouse IMX129840-2
18	72.5	5.4	196	21 AAY73344	HTM clone 0258181
19	71	5.3	159	19 AAW69170	N-terminally tagged
20	70.5	5.2	193	23 AAW76854	Human integrin alp
21	70.5	5.2	193	23 AAW76853	Human integrin alp
22	70	5.2	15	10 AAP98503	Sequence encoded b
23	70	5.2	170	22 AAU52472	Propionibacterium
24	70	5.2	202	22 AAG82686	S. epidermidis ope
25	69.5	5.2	170	22 AAU56330	Propionibacterium
26	69.5	5.2	184	21 AAB25446	Pinus radiata cell
27	69	5.1	134	21 AAG37220	Zea mays protein f
28	69	5.1	166	23 ABP26537	Streptococcus poly
29	69	5.1	194	22 AAG81764	S. epidermidis ope
30	68	5.1	180	22 AAG20206	Arabidopsis thalia
31	67.5	5.0	152	22 ABG19138	NOVEL human diageno
32	67.5	5.0	152	22 AAG37771	S. pneumoniae SGHR
33	67.5	5.0	170	19 AAB38722	Human ORFX ORF1346
34	67.5	5.0	177	21 AAB41582	Human MK61 protein
35	67.5	5.0	187	22 AAG91704	C glutamicum prote
36	67	5.0	178	22 AAG91926	Propionibacterium
37	66.5	4.9	83	22 AAU59092	Dermatophagoides p
38	66	4.9	181	19 AAW72342	Synthetic house du
39	66	4.9	181	22 AAY50544	S. pneumoniae derl
40	66	4.9	181	22 AAW19147	Human alpha 2 C4 a
41	66	4.9	183	19 AAW85905	Human deltydogenas
42	66	4.9	202	24 AAE32764	Partial corn extra
43	65.5	4.9	50	17 AAW03591	
44	65.5	4.9	102	23 ABP34479	
45	65.5	4.9	136	21 AAY44229	

ALIGNMENTS

RESULT 1	
ID	AAAP60627 standard; Protein; 196 AA.
XX	
XX	AAAP60627;
AC	25-MAR-2003 (updated)
XX	17-JUN-1991 (first entry)
DT	
DE	Beta-urogastrone - beta-lactamase fusion protein from pUC2101.
XX	
XX	Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
KW	fusion protein; beta-lactamase.
XX	
OS	Synthetic.
XX	
EH	Key
FT	Protein
FT	Location/Qualifiers
FT	1..120
FT	/label= beta-lactamase
FT	Misc-difference 121..123
FT	/label= adaptor
FT	Protein
FT	124..196
FT	/label= Beta-urogastrone
XX	
XX	DE3523634-A.
PN	
XX	09-JAN-1986.
PD	
XX	02-JUL-1985; 85DE-3523634.
PF	
XX	02-JUL-1984; 84JP-0137691.
PR	
XX	

NOVEL human diageno
NOVEL human diageno
Beta-urogastrone -
Vtgs-beta-lactama
Propionibacterium
H. pylori ORF 29ep
Human zcyto24 prot
Mouse IMX129840-2
HTM clone 0258181
N-terminally tagged
Human integrin alp
Human integrin alp
Sequence encoded b
Propionibacterium
S. epidermidis ope
Propionibacterium
Pinus radiata cell
Zea mays protein f
Streptococcus poly
S. epidermidis ope
Arabidopsis thalia
NOVEL human diageno
NOVEL human diageno
S. pneumoniae SGHR
Human ORFX ORF1346
Human MK61 protein
C glutamicum prote
Propionibacterium
Dermatophagoides p
Synthetic house du
S. pneumoniae derl
Human alpha 2 C4 a
Human deltydogenas
Partial corn extra

Db 61 AG 62

RESULT 15

Q8VQ00

ID Q8VQ00 PRELIMINARY; PRT; 128 AA.

AC Q8VQ00;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE CTX-M type beta-lactamase (Fragment).

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-clinical isolate HD3;

RA Stuerenburg E., Feucht H., Laufs R.;

RT "A new CTX-M type beta-lactamase."

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF466148; ANL73983.1; -

DR InterPro: IPR001466; Beta_lactamase.

DR Pfam: PF00144; beta-lactamase; 1.

DR PRINTS: PR00118; BLACTAMASEA.

DR NON_TER 1

FT 128

SQ SEQUENCE 128 AA; 13878 MW; F2EB815BB4FF318A CRC64;

Query Match

Best Local Similarity 22.2%; Score 299; DB 2; Length 128;

Matches 59; Conservativity 48.4%; Pred. No. 1.4e-17;

Matches 59; Conservativity 19; Mismatches 44; Indels 0; Gaps 0;

QY 66 LGRRIHYSNDLYEVSPTKEKLTIDGATVRELCSNAITMSDNTAANLLITIGGPKELTA 125

DB 6 LNRQVEIKKSDLVNPNPLAEKHVNGTMSLAELSAALQYSDNVAMNKLIAHVGGPASYTA 65

QY 126 FLHNGDVHTRLDRWEPELNEAIPNDRDRTMPVAMATTLRKILTGELLTASRQQLDW 185

DB 66 FARQIGDETFRLDRRETTLNNTAIPGDPDRTISPRAMQTLRNLTGLKALSDSQRAQLVTW 125

QY 186 ME 187

DB 126 MK 127

Search completed: September 10, 2003, 12:29:28
Job time : 26.4286 secs

```

DR InterPro: IPR000871: Beta_lactamase_A.
DR Pfam: PF00144: beta_lactamase; 1.
DR PRINTS: PR00118: BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 139 AA; 15257 MW; E514247C682442AD CRC64;
SQ
Query Match 33.7%; Score 453; DB 2; Length 139;
Best Local Similarity 66.9%; Pred. No. 1.6e-30;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLLTGGPKKELTAFLLHMGDHYVRLDRWPELNEALPNDERDTMPVAMATTLRLKLT
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 10 HLLATVGGPAGCTAFLKQIGDNTVRLDRWETELNEALPCDARDTTPASMAATLRLKLT
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 171 GELLTTLASROQLIDWMEADKVAPLLRSALPAGWFIADKSGAGERSGIIAALGPDGKP
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 SGRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKSGAKRGANGIYVALLGPNKA
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 231 SRIVVIT 237
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 130 ERIVVIT 136

RESULT 12
O53553 PRELIMINARY; PRT; 67 AA.
ID Q33553
AC Q33553;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAV-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSM 129;
RA MEDLINE=96081517; Pubmed=8538491;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81098; AAB35839.2; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT SEQUENCE 67 AA; 7287 MW; 85FEC85B0363F0CB CRC64;
SQ
Query Match 25.1%; Score 338; DB 2; Length 67;
Best Local Similarity 98.5%; Pred. No. 2.8e-21;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 55 VLSTRDAQDEQLGRIRHSNDLYEVSPTVEKHITDGMVRELCSAITSMDNTANILL
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1 VLSRDAGQEQGLGRIRHSNDLYEVSPTVEKHITDGMVRELCSAITSMDNTANILL
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 TTIGGPK 121
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 61 TTIGGPK 67

RESULT 13
O8GDE5 PRELIMINARY; PRT; 95 AA.
AC Q8GDE5;

```

```

DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_Taxid=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Ames S.G.B.;
RA "SHV-39: a new extended-spectrum beta-lactamase found throughout
RT Chile."
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY150585; AAN77730.1; -.
FT NON_TER 1
FT SEQUENCE 95 AA; 10299 MW; 496837847670A13C CRC64;
SQ
Query Match 23.8%; Score 320; DB 2; Length 95;
Best Local Similarity 64.2%; Pred. No. 1.5e-19;
Matches 61; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

OY 147 AIPNDEDTMPVAMATTLRLKLTGELLTSLASROQLIDWMEADKVAAGPLRSALPAGWFI
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 ALPGDARDTTPASMAATLRLKLTSGRLSARSQROLQWVDDRVAGPLIRSVLPAGWFI
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 207 ADKSGAGERSGRIIAGLPDGKPSRIYVITGTS 241
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 61 ADKTCASRGANGIYVALLGPNKARIVVITLRS 95

RESULT 14
O9JUN58 PRELIMINARY; PRT; 62 AA.
ID Q9JUN58
AC Q9JUN58;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAV-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSM 129;
RA MEDLINE=96081517; Pubmed=8538491;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81099; AAB35840.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT SEQUENCE 62 AA; 7049 MW; 1806CA190241540E CRC64;
SQ
Query Match 23.5%; Score 316; DB 2; Length 62;
Best Local Similarity 98.4%; Pred. No. 1.8e-19;
Matches 61; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 132 DHVTRLDWRPELNEALPNDERDTMPVAMATTLRLKLTGELLTSLASROQLIDWMEADK
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 1 DHVTRLDWRPELNEALPNDERDTMPVAMATTLRLKLTGELLTSLASROQLIDWMEADK
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 192 AG 193
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

Query Match      33.94; Score 456; DB 2; Length 138;
Best Local Similarity 66.94; Pred. No. 9, 1e-31;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

111 NLLTTTIGSPRELTAFLNHNGDHYRFLRDWEPELNEALPNDERDTPMVAAATTLRLT 170
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
10 NLLATVSGAGLTAFLRQIGDNYRFLDWEETLEALPRDARDTPYPASMAATLRLT 69

171 GELLTASRQOLIDNWEADKVASGPLLRSLAPAGFIIADSGSGSKGIIITAAATGPGPK 230
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
70 SQRISASGSQLLLMNVDDRVASGPLLRSLVPRAGWIIADTKTASRGARGIATALLGPNKA 139

```

RESULT 11	Q9AMA3			
ID	Q9AMA3	PRELIMINARY	PRT	139 AA.
AC	Q9AMA3			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DD	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
UD	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	SHV beta-lactamase (Fragment).			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
NC	Enterobacteriaceae; Escherichia.			
NR	NCBI_TaxID=562;			
LN	[1]			
RF	SEQUENCE FROM N.A.			
RC	STRAIN=935;			
RA	Subramaniam G., Navaratnam P.;			
RI	SHV gene sequences from ceftazidime-resistant Escherichia coli			
RL	isolates from UMUC, Malaysia.;			
RU	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF37047; AAK07462.1; -.			
DR	HSSP; P14577; ISHV.			
DR	InterPro; IPR001466; Beta_lactamase.			

```

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftriaxime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327049; AK07464.1; -.
DR EMBL: AF327051; AK07466.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 6.2e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTIGPKKELTALHNMGDHVTRLDRWPELNELINDEEDTTPVAMATTLRKLLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNVTRLDRWETELNLPDGDADTTTPASMAATLRKLLT 69
OY 171 GELLTASRQQLIDMKVADKVAAGPLRLSALPAGWFIADKSGAGERSGRTIALGPDGRP 230
DB 70 SQRLSARSQROLQNMVDDRVAGPLRLSVLPAGWFIADTGASKRGARGIVALGPNKKA 129
OY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 6
O9AMA1 PRELIMINARY; PRT; 138 AA.
ID O9AMA1;
AC O9AMA1;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=18/25;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftriaxime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327050; AK07465.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 6.2e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

```

```

OY 111 NLLTTIGPKKELTALHNMGDHVTRLDRWPELNELINDEEDTTPVAMATTLRKLLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNVTRLDRWETELNLPDGDADTTTPASMAATLRKLLT 69
OY 171 GELLTASRQQLIDMKVADKVAAGPLRLSALPAGWFIADKSGAGERSGRTIALGPDGRP 230
DB 70 SQRLSARSQROLQNMVDDRVAGPLRLSVLPAGWFIADTGASKRGARGIVALGPNKKA 129
OY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 7
O9AMA2 PRELIMINARY; PRT; 139 AA.
ID O9AMA2;
AC O9AMA2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=918;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftriaxime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327048; AK07463.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.1%; Score 458; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 6.2e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTIGPKKELTALHNMGDHVTRLDRWPELNELINDEEDTTPVAMATTLRKLLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNVTRLDRWETELNLPDGDADTTTPASMAATLRKLLT 69
OY 171 GELLTASRQQLIDMKVADKVAAGPLRLSALPAGWFIADKSGAGERSGRTIALGPDGRP 230
DB 70 SQRLSARSQROLQNMVDDRVAGPLRLSVLPAGWFIADTGASKRGARGIVALGPNKKA 129
OY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 8
O9AMA0 PRELIMINARY; PRT; 138 AA.
ID O9AMA0;
AC O9AMA0;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 263
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 105

RESULT 2

O52330 PRELIMINARY; PRT; 105 AA.

AC 052330;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Tnl bla protein (Fragment).
 OS Escherichia coli.
 OC Plasmid RK2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=562;

SEQUENCE FROM N.A.

RP MEDLINE=90264294; PubMed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The k11-kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korc.";
 RL J. Bacteriol. 173:3040-3050(1990).
 DR EMBL: M32794; AAA26408.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASA.
 KW Plasmid.
 FT NON_TER

SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1,1e-36;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 2 AMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 263
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 105

RESULT 3

O8RTD8 PRELIMINARY; PRT; 145 AA.

AC O8RTD8;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 NCBI_TaxID=573;

SEQUENCE FROM N.A.

RP Palasubramaniam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftazidime-resistant Klebsiella pneumoniae.";
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta_lactamase; 1.
 FT NON_TER
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 36.7%; Score 493; DB 2; Length 145;
 Best Local Similarity 68.1%; Pred. No. 7.4e-34;
 Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

OY 100 AATMSDNTAANLLITITGPKELTAEFLHNNGDVTRLDREPELNAIPNDEHDTTPY 159
 DB 7 AAYMSDNTAANLLITITGPKELTAEFLHNNGDVTRLDREPELNAIPNDEHDTTPY 66
 OY 160 AMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 67 SMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 126
 OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 263
 DB 127 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 144

RESULT 4

O52639 PRELIMINARY; PRT; 102 AA.

AC 052639;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Beta-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC Plasmid PRO1600.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=287;

SEQUENCE FROM N.A.

RP TRANSPOSON-Tn1;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RT "Deletion and transposon mutagenesis and sequence analysis of the
 RT pRO1600 OriK region found in the broad-host-range plasmids of the pOF
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL: L22691; AAA98312.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASA.
 KW Plasmid.
 FT NON_TER

SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1,1e-33;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 2 AMATTIRKLTGELLITLASROQLIDMWEADRVAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 263
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKM 144

RESULT 5

O990F3 PRELIMINARY; PRT; 138 AA.

AC 0990F3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SPO2_30E_37S

Perfect score: 1345
Sequence: I HPELVKVKDAEDQLGARV.....TMDERNQIAEIGASLIKHW 263

Scoring table:
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organellar:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriophage:*
17: sp-archaeo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	39.0	105	2	052026 pseudomonas
2	524	39.0	105	2	052330 escherichia
3	493	36.7	145	2	08RPD8 O8R18 klebsiella
4	488	36.3	102	2	052639 pseudomonas
5	458	34.1	138	2	0990P3 O9943 escherichia
6	458	34.1	138	2	0990P3 O9943 escherichia
7	458	34.1	138	2	0990P3 O9943 escherichia
8	458	34.1	138	2	0990P3 O9943 escherichia
9	458	34.1	138	2	0990P3 O9943 escherichia
10	458	34.1	138	2	0990P3 O9943 escherichia
11	458	34.1	138	2	0990P3 O9943 escherichia
12	458	34.1	138	2	0990P3 O9943 escherichia
13	458	34.1	138	2	0990P3 O9943 escherichia
14	458	34.1	138	2	0990P3 O9943 escherichia
15	458	34.1	138	2	0990P3 O9943 escherichia
16	458	34.1	138	2	0990P3 O9943 escherichia

17	241	17.9	48	2	O9RLH0 O9rlh0 proteus mir
18	238	17.7	52	2	O9R4I2 O9r4i2 shigella fl
19	197	14.6	38	2	P97145 O97145 escherichia
20	169.5	12.6	134	2	O8VUL3 O8vul3 staphylococ
21	124.5	9.3	109	2	O53698 O53698 staphylococ
22	105	7.8	20	2	P97146 P97146 escherichia
23	87	6.5	100	2	O93505 O93505 staphylococ
24	82	6.1	68	2	O9XBJ2 O9xbj2 bacillus ce
25	79	5.9	188	2	O9ACM8 O9acm8 streptococ
26	77	5.7	181	16	O8D108 O8d108 yersinia pe
27	76.5	5.7	200	16	O97JEB O97jeb clostridium
28	76.5	5.7	202	16	O98520 O98520 rhizobium l
29	75.5	5.6	184	5	O45379 O45379 caenorhabdi
30	74.5	5.5	195	16	O9RIA0 O9ria0 streptomyc
31	74	5.5	153	5	O8S520 O8s520 clona lites
32	74	5.5	177	16	O9KEM7 O9kem7 bacillus ha
33	73	5.4	175	17	O8RTH8 O8rth8 pyrobaculum
34	73	5.4	201	17	O8TH84 O8th84 methanosarc
35	72.5	5.4	113	2	O49970 O49970 mycobacteri
36	72.5	5.4	145	2	O05984 O05984 staphylococ
37	71.5	5.3	152	16	O9RI91 O9ri91 streptomyc
38	71	5.3	131	2	O9X9H0 O9x9h0 yersinia en
39	71	5.3	169	17	O972W2 O972w2 sulfolobus
40	71	5.3	192	16	O9HW07 O9hw07 pseudomonas
41	71	5.3	203	4	O96B23 O96b23 homo sapien
42	70.5	5.2	131	16	O8YOL8 O8yol8 anaeroba sp
43	70.5	5.2	180	16	O9ZKK0 O9zkk0 halobacte
44	70	5.2	108	17	O9HRO3 O9hro3 halobacteri
45	70	5.2	150	10	O94IG4 O94ig4 nicotiana t

ALIGNMENTS

RESULT 1

ID 052026 PRELIMINARY: PRT: 105 AA.

DC 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DI 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Beta-lactamase (Fragment).

OS Pseudomonas aeruginosa.

OG Plasmid pRO1514.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95011664; PubMed=7926843;

RA West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;

RT *Construction of improved Escherichia-Pseudomonas shuttle vectors

RT derived from pUC18/19 and sequence of the region required for their

RT replication in Pseudomonas aeruginosa.;

RU Gene 148:81-86(1984).

RU EMBL: L30112; AAA66058.1; -

DR HSSP: P00810; IXP.

DR InterPro: IPR001466; Beta_lactamase.

DR InterPro: IPR000871; Beta_lactamase_A.

DR Pfam: PF00144; beta-lactamase; 1.

DR PRINTS: PR00118; BLACTAMASEA.

KW Plasmid.

FT NON_TER

SQ SEQUENCE 105 AA: 11229 MW: D2889AA073330557 CRC64;

QY 160 AAMATTKLITGELTLASROOLIDMWDKVAAGPLRSALPAGRTIADKSGAGGSGNG 219

DB 2 AAMATTKLITGELTLASROOLIDMWDKVAAGPLRSALPAGRTIADKSGAGGSGNG 61

THIS PAGE BLANK (USPTO)

DB 105 AAIICGCHGVGODLHGEILTKNTLADVATLELYLLKVLIP 144

RESULT 13
ID HBG_RABIT STANDARD: PRT: 146 AA.
AC P02099:
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Hemoglobin gamma chain (Beta-3).
GN HBG.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=82053017; PubMed=6271761;
RA Hardison R.C.;
RT "The nucleotide sequence of rabbit embryonic globin gene beta 3.";
RL J. Biol. Chem. 256:11780-11786(1981).
RM (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=89178632; PubMed=2486295;
RA Margot J.B., Demers G.W., Hardison R.C.;
RT "Complete nucleotide sequence of the rabbit beta-1-like globin gene cluster. Analysis of intergenic sequences and comparison with the human beta-1-like globin gene cluster.";
RL J. Mol. Biol. 205:15-40(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN.
CC -1- SUBUNIT: Heterotrimer of two alpha chains and two gamma chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18818; AAA02984.1; -
DR EMBL; V00883; CAA24252.1; -
DR PIR; A02417; HBRB3.
DR HSSP; P02100; 1A9W.
DR InterPro; IPR002337; Beta_haem.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; globin; 1.
DR PRINTS; PR00814; BETAHAE.
DR PROSITE; PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Erythrocyte; Embryo.
FT INTRAMET 0
FT SEQUENCE 146 AA; 16093 MW; EB8DC1C24DD2D82 CRC64;

Query Match 4.68; Score 62.5; DB 1; Length 146;
Best Local Similarity 30.88; Pred. No. 2.6e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

DB 3 FTAEKKAITSWK-----LVYVEDAGAEALGRLL-----LVY-PWTORFFDSFGNL 48

QY 35 FRSERFPKMSSTFKVLLCGAVLSRIDAGQGLGRIRHYSONDLVSPYTEKHLITDGMV 94

DB 95 RELCSAIIITMSDNTAANLITTTGGPEKELTAFLLHNGDHVTRLD 138

DB 49 SS--SSAIMGNPVKAKH-----GKRVLTAF-----GDAVKND 79

RESULT 14

TATB_HELPY
ID TATB_HELPY STANDARD: PRT: 160 AA.
AC 025700:
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatb homolog.
GN TATB OR HP1060.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / AFCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tombl J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khakhria H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Waidman J.W., Fujii C., Bowman C., Watney L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter pylori.";
RL Nature 388:539-547(1997).
CC -1- FUNCTION: Required for correct localization of precursor proteins bearing signal peptides with the twin arginine conserved motif S/T-R-R-X-F-L-R. This sec-independent pathway is termed Tat for Lys-arginine translocation system. This system mainly transports proteins with bound cofactors that require folding prior to export (By similarity).
CC -1- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC -1- SIMILARITY: Belongs to the tatB family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000613; AAD08110.1; -
DR PIR; D64652; D64652.
DR TIGR; HP1060; -
DR HAMAP; MF_00237; - 1.
DR InterPro; IPR003369; Mta_Hcf106.
DR InterPro; IPR003998; TATB.
DR Pfam; PF02416; Mta_Hcf106; 1.
DR PRINTS; PR01506; TATBPROTEIN.
DR TIGRFAMs; TIGR01410; tatb; 1.
KW Transport; Protein transport; Translocation; Transmembrane;
KW Inner membrane; Complete proteome.
FT TRANSMEM 1
FT SEQUENCE 160 AA; 18332 MW; 3BDA4234F7BABBD7 CRC64;

Query Match 4.68; Score 62.5; DB 1; Length 160;
Best Local Similarity 20.58; Pred. No. 2.9e+02;
Matches 27; Conservative 30; Mismatches 64; Indels 11; Gaps 5;

QY 35 FRSERFPKMSSTFKVLLCGAVLSRIDAGQGLGRIRHYSONDLVSPYTEKHLITD-- 90

DB 19 FLGPEKFPQAVVDVYVFRAVYKTLNDADOTLDKNEINIEIKELTLEFKLENNVESIK 78

QY 91 GMTVRELCSAIIITMSDNTAANL--LITTTGGPEKELTAFLLHNGDHVTRLDREPELNEAI 148

DB 79 GKVIELEDAKVT--KENEIKSIQIDLMQYOKSLFTNTIPNHLNEVSN-----EALNKEV 133

QY 149 PNDERDTMPVA 160

```

CC EMBL; AB001728; AAD35606.1; -
DR PIR; G72365; G72365.
DR HSSP; P31059; 1E94.
DR MEROPS; T01.006; -.
DR TIGR; TM0521; -.
DR HAMAP; MF_00248; -. 1.
DR InterPro; IPR001353; Protease, Protease.
DR Pfam; PF00227; Proteasome; 1.
KW Hydrolyase; Protease; Complete proteome.
FT ACT SITE 6
SQ SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 3e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

OY 117 ICGPELTAFLHNMGDHVTLRDWRPELENEAIPNDRDTPMVAMATTIRKLTGELLTTL 176
DB 42 LGEGRVLAQFAGSVADAMTLFDREFAKLR-----MGNLTG 78
OY 177 ASRQQLIDMEADKYAGPLRSALPAGWFIADK-----SGAGKSGKGT-----AAL 224
DB 79 AAVEIAKDM-RTDRV-----LR-RLERALLVADKENIFITSGNGE-----VIOPPDDAAAI 127
OY 225 GPDG 228
DB 128 GSGG 131

RESULT 12
MARR_ECOLI
ID MARR_ECOLI STANDARD; PRT; 144 AA.
AC P27245; P76882; P75582;
DT 16-OCT-1992 (Rel. 23; Created)
DT 15-SEP-2003 (Rel. 42; Last sequence update)
DE Multiple antibiotic resistance protein marr.
GN MARR OR SOXO OR CFXB OR INAR OR B1530.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN (1)
RX SEQUENCE FROM N.A.
RP MEDLINE-93186717; PubMed-8383113;
RA Cohen S.P., Haechler H., Levy S.B.;
RT "Genetic and functional analysis of the multiple antibiotic
RT resistance (marr) locus in Escherichia coli.";
RL J. Bacteriol. 175:1484-1492(1993).
RN (2)
RX SEQUENCE FROM N.A.
RP STRAIN-K12 / MG1655;
RC MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:12453-1474(1997).
RN (3)
RX SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashiwano K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakano S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivaundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 mln region on the linkage map.";
RL DNA Res. 3:363-377(1996).

```

```

RN [4]
RX CHARACTERIZATION.
RP MEDLINE-94110219; PubMed-8282690;
RA Ariza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;
RT "Repressor mutations in the marrAB operon that activate oxidative
RT stress genes and multiple antibiotic resistance in Escherichia
RT coli.";
RL J. Bacteriol. 176:143-148(1994).
RN [5]
RX CHARACTERIZATION.
RP MEDLINE-95286534; PubMed-7768850;
RA Secane A.S., Levy S.B.;
RT "Characterization of MARR, the repressor of the multiple antibiotic
RT resistance (marr) operon in Escherichia coli.";
RL J. Bacteriol. 177:3414-3419(1995).
RN [6]
RX MARRGENESIS.
RP MEDLINE-20223625; PubMed-10760140;
RA Aleksun M.N., Kim Y.S., Levy S.B.;
RT "Mutational analysis of MARR, the negative regulator of marrAB
RT expression in Escherichia coli, suggests the presence of two regions
RT required for DNA binding.";
RL Mol. Microbiol. 35:1394-1404(2000).
CC -1- FUNCTION: REPRESSOR OF THE MARRAB OPERON WHICH IS INVOLVED IN THE
CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS
CC GENES. BINDS TO THE MARR OPERATOR/PROMOTER SITE.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M96235; AAC16394.1; ALT_INIT.
DR EMBL; AE000250; AAC74603.1; ALT_INIT.
DR EMBL; D90795; BAA15212.1; -.
DR EMBL; D90796; BAA15220.1; -.
DR EMBL; D90797; BAA15232.1; -.
DR PIR; E64907; E64907.
DR PDB; 1UGS; 28-DEC-01.
DR Ecogene; EGI1435; marr.
DR InterPro; IPR000835; HTH_MARR.
DR Pfam; PF01047; MARR; 1.
DR PRINTS; PR00598; HTHMARR.
DR SMART; SM00347; HTH_MARR; 1.
DR PROSITE; PS01117; HTH_MARR_FAMILY; 1.
KW Transcription regulation; DNA-binding; Repressor;
KW Antibiotic resistance; Complete proteome; 3D-structure;
FT MUTAGEN 45
FT REGION II TRANSCRIPT.
FT MUTAGEN 77 77
FT REGION II TRANSCRIPTION OF THE
FT MUTAGEN 123 144
FT REGION II TRANSCRIPT.
FT REGION II TRANSCRIPTION OF THE
FT SEQUENCE 144 AA; 16065 MW; BE7DF5549E24DD3 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 144;
Best Local Similarity 29.4%; Pred. No. 2.5e+02;
Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

OY 66 LGRRIHY--SQND--LVEY--SP--VTEKLTGMYRELCSAITSNTAANLLTTTG 118
DB 14 LGRLLHVNQKQRLNELYISPLDITAAQFYKCSIR--CAACIT----- 56
OY 119 GPELTAFLH-INGCHVTGLDR-----WEPELEAIPNDRDTPMVAMATTIRKLTG- 171
DB 57 -PVEIAKRVSDVAGALTRLDRLVCKGWERLPN--PDKRG-----VLVLTG 104
OY 172 -----ELTLASRQQLIDMEADKYA--GPLRSALP 201

```


CC EMBL: AF001079; AAB90876.1; -
 CC PIR: C69296; C69296.
 DR TIGR: AF0371; -
 DR HAMAP: MF_00241; -; 1.
 DR InterPro: IPR003265; Endo_3c.
 DR Pfam: PF00730; HHH-GPD; 1.
 DR SMART: SM00478; ENDO3C; 1.
 DR SMART: SM00478; ENDO3C; 1.
 DR MultiFunctional enzyme; Endonuclease; Lyase; DNA repair; Glycosidase;
 KW Multicatalytic enzyme; Complete proteome.
 FT ACT_SITE 122 122 BY SIMILARITY.
 SQ SEQUENCE 198 AA; 22639 MW; 3A5C033AA12FFFB CRC64;

Query Match 4.8%; Score 65; DB 1; Length 198;
 Best Local Similarity 24.6%; Pred. No. 2.4e+02;
 Matches 51; Conservative 25; Mismatches 85; Indels 46; Gaps 14;

OY 14 QLGARVGYIELDN---SGEILIESFRSEERPMSTFEVLICGAVLSRIDGQ-----EQ 65
 DB 15 QLGKRGVEFDREFRDLFSKATITRTLAFICISTANSSAVAGLKFQRL-LEQGVGVK 72
 OY 66 L---GRRIHYSQNDLVEYSPYTERHLDGKMTVREICSAATMSDNTAANLLTTIG-GPK 121
 DB 73 LTLGVRFH---NKRKAEY-----IREAFKSPFLYERALEASSKAREILKTKIGLGM 122
 OY 122 ELTAFPLHMG-DHYTRLD---RWEPELNEAIPNDERDTPVPVAMATTLRKL-----T 170
 DB 123 EASHFLRWGREDVAIIDRLHMLEROGYEVPG---TWTAKRYLEVERKILMEISEER 177
 OY 171 GELLTLASROOLIDMMEADKVPPLR 197
 DB 178 GESL---AEMDLRIWAE---WTGKVLK 198

RESULT 10

RR4_PELINE STANDARD; PRT; 200 AA.
 AC Q9FSD9;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chloroplast 30S ribosomal protein S4.

GN RPS4.
 OS Pellia neesiana (Liverwort).

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
 OC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellilnaceae;

OC Pellilnaceae; Pellia.
 OX NCBI_Taxid=70144;

RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Gametophyte;

RA Capesius I., Bloecher R.
 RT "A molecular approach to bryophyte systematics."

CC Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: One of the primary rRNA binding proteins, it binds

CC directly to 16S rRNA where it nucleates assembly of the body of
 CC the 30S subunit (by similarity).

CC -1- FUNCTION: With S5 and S12 plays an important role in translational
 CC accuracy (by similarity).

CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
 CC The interaction surface between S4 and S5 is involved in control

CC of translational fidelity (by similarity).
 CC -1- SUBCELLULAR LOCATION: Chloroplast.

CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib-sib.ch).

CC EMBL: AJ250456; CAC14050.1; -
 CC HSSP: P81288; IC05.
 DR HAMAP: MF_01306; -; 1.
 DR InterPro: IPR001912; Ribosomal_S4.
 DR InterPro: IPR002942; S4.
 DR InterPro: IPR005709; S4_bact_Org.
 DR Pfam: PF00163; Ribosomal_S4; 1.
 DR Pfam: PF01479; S4; 1.
 DR SMART: SM00363; S4; 1.
 DR TIGRFAMs: TIGR01017; rpsd_bact; 1.
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE: PS50889; S4; 1.
 KW Ribosomal protein; RNA-binding; RNA-binding; Chloroplast.
 FT DOMAIN 90 150 S4 RNA-BINDING.
 SQ SEQUENCE 200 AA; 23178 MW; 00D0098C6D933F5 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 200;
 Best Local Similarity 22.8%; Pred. No. 3.2e+02;
 Matches 39; Conservative 31; Mismatches 78; Indels 23; Gaps 7;

OY 14 QLGARVGYIE--LDNSGEILIESFRSEERPMSTFEVLICGAVLSRIDGQPOLGRH 71
 DB 14 RLGLPGLTFSKILELESGYIGOSTPNKK---VSQYRRL-----EKOKLRH 58
 OY 72 Y---SONDLYEYSPYTERHLDGKMTVREICSAATMSDNTAANL-LTTIGGPRELTAFLH 128
 DB 59 YGLEROLLYVRIARK--AKGSTGQTLISLTLERKLDNIFIRGMSPTTIGARQLVNHHR 116
 OY 129 NM-GDHYTRLDRWEPELNEAIPNDERDTPVPVAMATTLRKLGGELLTAS 178
 DB 117 ILINDNTVDIPSYNCEPKDYITVNNRRESYIIINMDSKRPKPNHLITDS 167

RESULT 11

HSLV_THEME STANDARD; PRT; 176 AA.
 AC Q9WY21;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-dependent protease hslV (EC 3.4.25.-).

GN HSLV OR TM0521.

OS Thermotoga maritima.

OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_Taxid=2336;

RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;

RA MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Uitterback T.R., Malek J.A., Liner K.D., Garrett K.M.,

RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,

RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
 RT Evidence for lateral gene transfer between Archaea and Bacteria from

RT genome sequence of *Thermotoga maritima*.
 RL Nature 399:323-329(1999).

CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (by similarity).

CC -1- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUPFAMILY
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib-sib.ch).

DR HAMAP: MF_00003; -; 1.
 DR InterPro: IPR000238; Rib_bind_facta.
 DR Pfam: PF02033; RBFA; 1
 DR Pfam: PF007327; Rib_bind_facta; 1.
 DR TIGRFAMs: TIGR00082; rDNA; 1.
 DR PROSITE: PS01319; RBFA; FALSE_NEG.
 DR RNA PROCESSING: Complete proteome.
 KM SEQUENCE 116 AA; 13247 MW; A79DCG71F0547514 CRC64;
 SO
 Query Match 4.9%; Score 65.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No. 1.1e+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;
 OY 144 LNEAIPDERDPTVAMATTLRLKLTGEL-----LTIASRQOLIDME-ADKVAAPLL 196
 DB 18 INNALANETNDKINKLAVTAVR--LSNDLSVAKFFLDHAKRESMLKYLEVNNKVS-LL 74
 OY 197 RSALPAGW-----FIADKS 210
 DB 75 RSKLAEWTSYKVPRLRFVIDET 97
 RESULT 8
 KTHY_METAC STANDARD; PRT; 206 AA.
 ID KTHY_METAC
 AC O8THS9;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN TMK OR MA4433.
 OS Methanosarcina acetivorans.
 OC Archaea: Euryarchaeota: Euryarchaeota orders incertae sedis;
 OC Methanosarcinales: Methanosarcinaceae: Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nuebaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., Dekrellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuelner H.C., Krzycki J.A.,
 RA Leigh J.A., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umeyan L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:533-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
 CC 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE011164; AM07774.1; -;
 DR HAMAP: MF_001165; -; 1.
 DR InterPro: IPR000062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1.
 DR TIGRFAMs: TIGR00041; dTMP_kinase; 1.
 DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
 DR Transferase: kinase; Nucleotide biosynthesis; ATP-binding;
 KM Complete proteome.

FT NP_BIND 10 17 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 23303 MW; A19C2E5BC85423B3 CRC64;
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.3e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 OY 56 LSRID-AGOEOLGRIRIHSYNDLYEVSPTKEKLTGDMTVELCSAAITMSDNTANLL 114
 DB 8 LEGIDSGKSTVANKLQ-ENSELRFEPVETREPTRG-----TLGNVAVENAIQ 55
 OY 115 TTIGGPPLETFLLNMGDYRLDRWPELNEAIPNDEROTVPMVAMATTLRLKLTGEL 174
 DB 56 SDTQLAELFTADHAEHLAKLVKPALEDKTVISDRYSDSRVAYOGITLKNRLDNPV- 114
 OY 175 TLASRQOLIDMEADKVAAPLLRSALPAGW-----FIAD-----KSG-AGE----- 214
 DB 115 -----EWVR-----DIARGWVVIDLFLFDIEELAVKRCGRKGEQTKF 154
 OY 215 -----RGSRCIIIALGPDGKPSRIVAYTTGSOATYDE 247
 DB 155 EKIEFLRGVRELFLGLAAE-EPERFVIYDASGSDPEVEK 192
 RESULT 9
 OG1_ARCFU STANDARD; PRT; 198 AA.
 ID OG1_ARCFU
 AC 029876;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable N-glycosylase/DNA lyase (includes: 8-oxoguanine DNA
 DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
 DE (EC 4.2.99.18) (AP lyase)].
 GN OGG OR AF0371.
 OS Archaeoglobus fulgidus.
 OC Archaea: Euryarchaeota: Archaeoglobi: Archaeoglobales;
 OC Archaeoglobaceae: Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=93896475;
 RA Klek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.R., Peterson J.D.,
 RA Richardson D.D., Kerlavage A.R., Graham D.E., Kyrylski N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artlich P., Kalne B.P., Sykes S.W.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997).
 CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
 CC guanine (7,8-dihydro-8-oxoguanine = 8-oxoG) from DNA. Also nicks
 CC DNA at apurinic/apyrimidinic sites (AP sites) (BY similarity).
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
 CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
 CC leaving a 3'-terminal unsaturated sugar and a product with a
 CC terminal 5'-phosphate.
 CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; LA7838; AAB38480.1; -.
DR EMBL; Z99115; CAB14127.1; -.
DR PIR; H69647; H69647.
DR Subtilisin; BG11396; kgdA.
DR InterPro; IPR000887; Aldolase_KDPG_KMG.
DR Pfam; PF01081; Aldolase; 1.
DR TIGRfam; TIGR01182; eda; 1.
DR PROSITE; PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
DR Kyr; PS00160; ALDOLASE_KDPG_KMG_2; 1.
DR Lyase; Schiff base; Multifunctional enzyme: Complete proteome.
FT ACT SITE 43 43 BY SIMILARITY.
FT ACT SITE 47 47 BY SIMILARITY.
FT BINDING 130 130 SCHIFF-BASE WITH KMG OR PYRUVATE (BY
FT BINDING 130 130 SIMILARITY).
SQ SEQUENCE 196 AA; 20865 MW; EB08597FD1102CC4 CRC64;

Query Match 4.9%; Score 66.5; DB 1; Length 196;
Best Local Similarity 21.6%; Pred. No. 1.8e+02;
Matches 27; Conservative 19; Mismatches 60; Indels 19; Gaps 4;

QY 28 SEILSEFRSEERFPMSTFKVLLCGAVLSRIDAG-EOIGRIHYSON---DLVEYSPV 83
DB 51 ASDIIEFRRNREDI-----LIGACTVISAQOAGAAKAGQFVSPFSADLAHLSF 103
QY 84 TEKHLDGMYRELCSAATMSDNTAANL-----LTTTGPKELTAFLHNGDHYT 135
DB 104 VTHHYGVLTPESEMEALTFGFTYLLKLPFGVFGIPKMKLAGPPOVTFIPTGIGHPS 163
QY 136 RLDDP 140
DB 164 EYDPW 168

RESULT 6
NDK_TREPA STANDARD; PRT; 149 AA.
ID NDK_TREPA
AC 083974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
DE (Nucleoside 2'-p-kinase).
GN NDK OR TP1010.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Doderger E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ulteback J.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
CC OTHER THAN ATP.
CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate -> ADP +
CC nucleoside triphosphate.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the NDK family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001269; AAC65961.1; -.
DR PIR; F71252; F71252.
DR HSP; P22887; INRK.
DR TIGR; TP1010; -.
DR HAMAP; MF_00451; -.
DR InterPro; IPR001564; NDK.
DR Pfam; PF00334; NDK; 1.
DR PRINTS; PR01243; NUCOPKINASE.
DR PRODOM; PD001018; NDK; 1.
DR SMART; SM00562; NDK; 1.
DR PROSITE; PS00469; NDP_KINASES; 1.
DR Transferrase; Kinase; ATP-binding; Complete proteome.
FT ACT SITE 117 117 BY SIMILARITY.
SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

Query Match 4.9%; Score 66; DB 1; Length 149;
Best Local Similarity 21.6%; Pred. No. 1.4e+02;
Matches 35; Conservative 24; Mismatches 65; Indels 38; Gaps 7;

QY 29 GEILSEFRSEERFPMSTFKVLLCGAVLSRIDAGEOIGRIHYSONDLVEYSPVTEKHL 88
DB 21 GEVLSRF---ERKGVLTALRLC-----VDIATAEI---HYAHEKRPYSLIAYI 67
QY 89 TDGMYRELCSAATMSDNTAANLTTTGPKELTAFLHNGDHYTLRDPWELNEAI 148
DB 68 TSAIPV-----ALAFKGENAISLVRLCGS-----TRVEIAQP---GTI 103
QY 149 PND-ERDTTPVAMATLRLKLTGELLTLASROOLIMWEND 189
DB 104 RGFALRTTIVIVASDSPEASAREALALYSAQDFVEMRGN 145

RESULT 7
RBEA_UREPA STANDARD; PRT; 116 AA.
ID RBEA_UREPA
AC 09FOH0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBEA OR UO321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sevovar 3;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caspell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002129; AAF30730.1; -.

```

```

QY 44 MSTEKVLGCVNLSDIAGQEQGLGRIRHYSNDLY-EXP-VTEKHLTDGMTRELCSNA 101
  1 MMSIVVCVVG-----EGIGTSMLSLTYSNSISNETQPYFDVYST----- 42
DB 102 IYMSDNFTANLLTTTIGCKELTAP-----LANNGDVTRLDRNEPELN 145
QY 43 LLMHKKRPYNLSLMTPTAGGEFSKLRSLSPQTDVFLCFLSLINPSSSNILDSVWQELN 102
DB 146 EAPIDDERDTMPVAMATTLRKLLTGELL--TLASRQOL 182
QY 103 ENCPN-----TPYLVGTQMDLKSNSVTLDRLCERKOL 135
DB

RESULT 4
HC3L_THIFE STANDARD; PRT; 122 AA.
AC P80509;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome-C3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
DE (Fragmenta).
GN HOXG.
OS Thiobacillus ferrooxidans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
OC Acidithiobacillaceae; Acidithiobacillus.
OC NCBI_TaxId=920;
OX
RN
RP SEQUENCE.
RC STRAIN=ATCC 19859;
RA MEDLINE=96241862; PubMed=8661919;
RA Fischer J., Quentmeyer A., Kostka S., Kraft R., Friedrich C.G.;
RT "Purification and characterization of the hydrogenase from
RT Thiobacillus ferrooxidans."
RL Atch. Microbio. 165:289-296(1996).
CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferriocytochrome c3 = 2 H(+) + 2
CC ferriocytochrome c3.
CC -1- COFACTOR: IRON.
CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
KM Oxidoreductase; Iron
FT NON_CONS 20 21
FT NON_CONS 29 30
FT NON_CONS 35 36
FT NON_CONS 42 43
FT NON_CONS 59 60
FT NON_CONS 72 73
FT NON_CONS 78 79
FT NON_CONS 87 88
FT NON_CONS 98 99
FT NON_CONS 107 108
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 13146 MW; EB4FAF365348DA0B CRC64;

Query Match 5.18; Score 68; DB 1; Length 122;
Best Local Similarity 26.48; Pred. NO. 74;
Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

QY 19 VGYIELDNGSEILSEFRSERPFMSTFVLLCGAVLSRIDAQEQGLGRIRHYSNDLY 78
DB 12 VGRREGDLSKSTLEFRN-----ALNRPGGG-----LG 41
QY 79 EYSPVTEKHLTDGMTRELCSNAITMSDNFTANLLTTTIGCKELT 124
DB 42 KYAPFTGTNVEIGVT-----ISGDKDPQAGLVT-----PREST 75
DB

RESULT 5
ALKH_BACSU STANDARD; PRT; 196 AA.
ID ALKH_BACSU
AC P50846;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE KHG/KDGC aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase
DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KHG-aldolase); 2-
DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
DE dehydro-3-deoxygluconate aldolase) (phospho-2-keto-3-deoxygluconate
DE aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDGC-
DE aldolase)].
GN KDCG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OC NCBI_TaxId=1423;
OX
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Madrid;
RA MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kdg loci cloned in a yeast artificial chromosome."
RL Microbiology 142:2005-2016(1996).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chol S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppl G., Guy B.J., Haga K., Halesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchlyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viaro A., Wambolt R., Wedler E., Wedler H., Welfzenegger T.,
RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate - pyruvate +
CC glyoxylate.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE KHG/KDGC ALDOLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E_37S
Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARVG.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78	5.8	191	YDA4_SCHPO	Q10346 schizosacch
2	69.5	5.2	196	RAC2_LOCTA	Q40220 lotus japon
3	69	5.1	201	RACG_DICDI	Q9980 dicystostell
4	68	5.1	122	HCB3L_THFE	P80309 thlobacilli
5	66.5	4.9	196	ALKH_BACSV	P50846 b khg/dcp
6	66	4.9	149	NDK_TREPA	O83974 treponema p
7	65.5	4.9	116	RBR4_UREPA	Q99qho ureaplasma
8	65.5	4.9	206	KTHY_METAC	Q8bts9 methanosarc
9	65	4.8	198	OGGI_ARCFU	Q23676 archaeeoglob
10	63.5	4.7	200	RR4_PELNE	Q9f6d9 pelilla nees
11	63	4.7	176	HSIV_THEMA	Q9wyz1 thermotoga
12	62.5	4.6	144	MARR_ECOLI	P27245 escherichia
13	62.5	4.6	146	HBC_RABIT	P02099 cyctolagus
14	62.5	4.6	160	TATB_HELPY	Q25700 helicobacte
15	62.5	4.6	172	YDEJ_ECOLI	P31131 escherichia
16	62.5	4.6	182	PYRE_STRCO	Q9x8r7 streptomyce
17	62.5	4.6	184	HRPL_PSESY	P37929 pseudomonas
18	62.5	4.6	195	TRPF_THEVO	Q979v6 thecoplasma
19	62	4.6	173	Y265_BORBU	O51280 borrelia bu
20	62	4.6	174	IRHB_RAT	P17991 rattus norv
21	62	4.6	182	RR4_BEJCH	O19990 belamcanda
22	62	4.6	182	Y861_BIFLO	Q89320 bifidobacte
23	61.5	4.6	152	Y033_YERPE	Q8zcf8 yersinia pe
24	61.5	4.6	184	MLR1_SCHPO	Q9uug5 schizosacch
25	61.5	4.6	185	YCJC_ECOLI	P38322 escherichia
26	61.5	4.6	197	HANL_XENLA	Q73815 xenopus lae
27	61	4.5	177	VNSC_RINDK	P35348 rinderpest
28	61	4.5	177	VNSC_RINDK	Q06339 rinderpest
29	61	4.5	184	ARL2_DROME	Q06649 drosophila
30	61	4.5	194	RR4_TRIEN	O20234 tris ensata
31	60.5	4.5	126	YGM1_YEAST	P53130 saccharomyc
32	60.5	4.5	142	YE63_METTM	O50770 methanobact
33	60.5	4.5	150	PDUV_SALTU	Q9xdm6 salmonella

RESULT 1	YDA4_SCHPO	STANDARD:	PRT:	191 AA.
AC	Q10346:			
DT	01-OCT-1996 (Rel. 34, last sequence update)			
DT	28-FEB-2003 (Rel. 41, last annotation update)			
DE	Hypothetical protein Clp12.04c in chromosome 1.			
GN	SPAC1F12.04c.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomyces.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=972;			
RX	MEDLINE=21848401: PubMed=11859360;			
RA	Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,			
RA	Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,			
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,			
RA	Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,			
RA	Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,			
RA	Holtrold S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,			
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,			
RA	Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,			
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,			
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,			
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,			
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,			
RA	Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,			
RA	Welfjens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,			
RA	Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,			
RA	Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,			
RA	Eger P., Zimmermann W., Wedler H., Mambert R., Purrelle B.,			
RA	Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motter S.,			
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurt S.M.,			
RA	Lucas M., Rochet M., Galliard C., Tallada V.A., Garon A., Rhode G.,			
RA	Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,			
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,			
RA	Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,			
RA	Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;			
RT	*The genome sequence of Schizosaccharomyces pombe.*;			
RL	Nature 415:871-880(2002).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: Z69944; CAA93808.1. -			
DR	PIR: T38062; S67447.			
DR	GeneDB_Spombe: SPAC1F12.04c; -			

THIS PAGE BLANK (USPTO)

AD2281
 hypothetical protein alr3803 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AD2281
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yaeu, M.; Tabeta, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. strain PCC 7120
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AD2281
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1131 <OR>
 A:Cross-references: GB:BA000019; PIDN:BAW5502.1; PID:g1732937; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr3803

Query Match 5.2%; Score 70.5; DB 2; Length 131;
 Best Local Similarity 26.0%; Pred. No. 81;
 Matches 34; Conservative 19; Mismatches 51; Indels 27; Gaps 5;

OY 77 LVEYSPYTEKHLTDGTVRELCSAA-TTMSDNTAANLLTTGGPKELTAPLHNM--GDH 133
 DB 1 MWSSSVTLQNTDNLVLTFLTFSSQNCQLYDQTEPLTLDPPYYSKERNFTTALYAGNF 60
 OY 134 VTRLD--RKEPELNAIPNDERDTMPVAMATTLKRLT-----GE 172
 DB 61 VTSPFMTWODEANRFTVNPPE---LLNLADISTICKLTSVRKRRCFSGHQAQMDNGH 117
 OY 173 LTTLASROOLI 183
 DB 118 LFTIQRQAI 128

RESULT 11
 C71869
 hypothetical protein jhp0935 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: C71869
 R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; Kling, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: C71869
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-180 <ARN>
 A:Cross-references: GB:AE001523; GB:AE001439; NID:94155513; PIDN:AD06513.1; PID:9415551
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0935

Query Match 5.2%; Score 70.5; DB 2; Length 180;
 Best Local Similarity 25.4%; Pred. No. 1.2e+02;
 Matches 47; Conservative 26; Mismatches 43; Indels 67; Gaps 12;

OY 30 EILSFSEBEPFPMSTFKVLLCGAVLSRIDAGQEQGRRIHYSQNDLV-----EYSPYTE 85
 DB 2 ETFAITAEKEREPFSLF-----NRSSGSDDLTKQVSKYENTLIDTKEYSREYO 52
 OY 86 KHLTDGTVRELCSAATMSDNTAANLL--TTIGGPKEL-TAFLLHMGDHTRLDMEP 142
 DB 53 K-----ANLIS-----NIVLVPPT--PSQIDTEVLNMLERIQIQ----- 86
 OY 143 ELNE--AIPNDERDTMPVAMATTLKRLTGTLLTASROOLIIMME--ADXYAGPL 195
 DB 87 ELNENELALIVINRMPITP-----TLKEQALIFETIKENPSDRIT--L 128

OY 196 LRSAL 200
 DB 129 LESSL 133

RESULT 12
 E84217
 hypothetical protein Vng0594h [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84217
 R:Ng, W.Y.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Lethausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja, Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Title: Genome sequence of Halobacterium species NRC-1
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: E84217
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA619105.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: Vng0594h

Query Match 5.2%; Score 70; DB 2; Length 108;
 Best Local Similarity 27.5%; Pred. No. 68;
 Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

OY 44 MSTFKVLLCGAVLSRIDAGQEQGRRIHYSQNDLVESPYTEKHLTDGTVRELCSAAT 103
 DB 1 MDVTVELDQALLARLRTD-----RFEVEFPALTEVTTLRFHGDGRV-----GSIV 50
 OY 104 MSDNTAANLLTTIGGPKEL-----TAFLLHMGDHTRLDR 139
 DB 51 NDDGTDRMTARLITVPGSDFIAVEVPTSFVAALVDATRI DR 92

RESULT 13
 F84251
 hypothetical protein Vng0953c [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84251
 R:Ng, W.Y.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Lethausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja, Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Title: Genome sequence of Halobacterium species NRC-1
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: F84251
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <STO>
 A:Cross-references: GB:AE004437; NID:g10580510; PIDN:AA619378.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: Vng0953c

Query Match 5.2%; Score 69.5; DB 2; Length 145;
 Best Local Similarity 20.3%; Pred. No. 1.1e+02;
 Matches 30; Conservative 29; Mismatches 66; Indels 23; Gaps 3;

OY 41 FPMSTFKVLLCGAVLS-----RIDAGQEQGRRIHYSQNDLVESPYTEKH 87
 DB 3 FSVSGSAALITFAATVSGILYSAAFNGFERQDANNAANDRYLAKTVEVANTYDS 62
 OY 88 LFDGTVRELCSAATMSDNTAANLLTTIGGPKELTAFLLHMGDHTRLDMEP 147
 DB 63 VNDGTVAATNNGSTLISVSTQTDVY---DEFTYDSATVSSVDGNSQTDLM----- 112

C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: A83753
 R:Takami, H.; Nakasono, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: A83753
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-177 <STO>
 A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04544.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0825
 C:Superfamily: *Bacillus subtilis* hypothetical protein ykka

Query Match 5.5%; Score 74; DB 2; Length 177;
 Best Local Similarity 20.3%; Pred. No. 62;
 Matches 38; Conservative 29; Mismatches 56; Indels 64; Gaps 8;

QY 42 PMNSTFKVLLCGAVLSRIDAGEQLGRIRHYSQNDLVESPYTEKHLTDGKTVRELCSAA 101
 DB 11 PEMDSVGLFYMAYEENVND-----RLH-----HLIEDVTEELYYKG 47
 QY 102 ITMSDNTANLLTTIGGPKELTAFLNMGDHTBLD-RWEDEL-NEAIPND----- 151
 DB 48 SDGDSMAQLL-----NHLITVAMVVRIGFALPDSLBAHGM 89
 QY 152 -ERDTTPYAMATTLTKLITGELLTLASRQ-----LIDMEADKVAQPLRSALP 201
 DB 90 VDKDKLPVTVLSVGLIEKORYVALLEKTCQALHDDDLARWIPYEERQATIRWGL- 148
 QY 202 AGCFIND 208
 DB 149 --WHMAD 153

RESULT 7
 T45195
 hypothetical protein u1756c [imported] - *Mycobacterium lepre*
 C:Species: *Mycobacterium lepre*
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T45195
 R:Robison, K.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: Z16911
 A:Accession: T45195
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-113 <KEI>
 A:Cross-references: EMBL:U015180; PIDN:AAA62885.1

Query Match 5.4%; Score 72.5; DB 2; Length 113;
 Best Local Similarity 28.6%; Pred. No. 45;
 Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;

QY 130 MGDHV--TRLDWEPELN--EAIPINDERDT--TPVAMATTLTKLITGELLTLASRQ 180
 DB 1 MGDGIGEREGKRTGNTGCPRLAVPGDESPITDGRASPEDLIIT--NLSPPTMSHPPS 58
 QY 181 QLDWMEA-DKVAPELL-----RSALPAGWFTADKSGAGERSGRTIAALGPDGKFSR 232
 DB 59 RDDDWVPEPDLQCTAVFDATGDKATPAVAGI---GASTRGS-GILASISPEFROPAR 112

RESULT 8
 T36984
 hypothetical protein SCU11.13 - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
 C:Accession: T36984
 R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1999
 A:Reference number: Z21618

A:Accession: T36984
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-152 <OL1>
 A:Cross-references: EMBL:AL109949; PIDN:CBAS2898.1; GSPDB:GN00070; SCOEDB:SCU11.13
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCU11.13
 C:Superfamily: *Streptomyces coelicolor* hypothetical protein SCU9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
 Best Local Similarity 23.4%; Pred. No. 81;
 Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

QY 106 DNTANLLTTIGGPKELTAFLNMGDHYRDLDEPELNRAIPNDERDTTPVAMATTL 165
 DB 25 DQAAKRTLAQYTPPADLATEYESIGD-VT---WEDVNGNYFLNP----- 65
 QY 166 RKLLTGELLTLASROOLIDMEADKVAQPLRSALPAGWFTADKSGAGERSGRTIAALG 225
 DB 66 ----AGDLLAKQEVGVDPFGTDEKSRGLV-----GSNGG-GLIYVAG 104
 QY 226 PDGKPSRIVITYTGSQATMDERN-----RQIAET 255
 DB 105 PDG-----AVYRTRT-ASLDEALDKVADGIRQPLEL 135

RESULT 9
 G83096
 conserved hypothetical protein PA4399 [imported] - *Pseudomonas aeruginosa* (strain PAO
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
 C:Accession: G83096
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
 Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Ladbeg, K.; L
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: G83096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <STO>
 A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA07787.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4399
 C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
 Best Local Similarity 26.0%; Pred. No. 1,2e+02;
 Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

QY 53 GAVLSRI-----DAGEOL--GRRIRHYSQNDLVESPYTEKHLTDGKTVRELCSA----- 100
 DB 2 GNLRSITVTRTGGTGLAGGRVPSRPIAICAVDELNSQDLLAELLEANGAMP 61
 QY 101 -----ATMSDNTANLLTTIGGPKELTAFLNMGDHYRDL--DWEPEL-----N 145
 DB 62 GLEIYQALAPVGR-----LFDGGLAMPETVRALDEVRALDESCIDRWMDLGLPLKN 116
 QY 146 EAIPINDERDTTPVAMATTLTKLITGELLTLASROOLIDMEADKVAQPLRSALPAGW 205
 DB 117 FILPGGSR---PVAQHVCRLAR---SAERCCALDQEEITLGVGLRYLRLDLLE 168
 QY 206 IADKSGAGERSGRTI--AALGPD 227
 DB 169 VAAARAIKRGVAVETIEMAAKPD 192

RESULT 10

Best Local Similarity 22.6%; Pred. No. 32;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;

OY 104 MSDNTAANLTLTTIGPKELTAFLHNMGDHVTLRDWEPELNAIRNDRDTPMAMAT 163
DB 1 MSYHSLNIMLQPSGIDRIKAIILVN-----VARLD-----PASKSTHQLVSMIN 46
OY 164 TLRLKLLTGLTLTLASQQLIDMWENDKVAGPLLRSLALPAGWETADKSGAGERSGRIIA 223
DB 47 ERFRC-----ILRLPLGLKLIIVNFRKDSPEFTYMSNAINIGYVTE--GLAFLGKQIISI 99
OY 224 LSPDGRP-----SRIYVYTTGSOATMDERNQI----- 252
DB 100 -----SKPLEDXLMWSSRWLDTLTITQLRRETEDEKHEQDLASNLASLPLCTIHS 155
OY 253 AEIGASLRKH 262
DB 156 VENGAGLHKH 165

RESULT 3

G97064

Spore coat protein COTUC [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: G97064

R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J. Bacteriol. 183, 4833-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CLO

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: G97064

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-200 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:915024270; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1338

Query Match 5.7%; Score 76.5; DB 2; Length 200;
Best Local Similarity 25.7%; Pred. No. 45;

Matches 36; Conservative 26; Mismatches 55; Indels 23; Gaps 7;

OY 1 HPEFLVAVKDAEDQAGRVGYIELDLSNGEILLESFR-SEERFPWMTFFKVLGCAYLSRI 59
DB 10 HP---VRIKPNNAQL-AKVITTYGPGDGLAASIRYLSQFSNVT-----QAIATLN 59
OY 60 DAGEOIGRIRHYSONDLVEPSVTEKHLTDGMTVRELCSNAIT--MSDNTAANLTLTTI 117
DB 60 DIGEEELAH-----LEIVGSIVROLRSGLSVEELKRSGLDAPADHDSALYPSAA 110
OY 118 GCPKELETAFLHNMGDHVTRL 137
DB 111 GNP-FTAAVYQSGKDPITDL 129

RESULT 4

T21126

ADP-ribosylation factor homolog F19H8.3 [similarity] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001

C:Accession: T21126

R:Steward, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19379

A:Accession: T21126

A:Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-184 <WTI>

A:Cross-references: EMBL:Z93378; PIDN:CAB07583.1; GSPDB:GN00020; CESP:F19H8.3

A:Experimental source: clone F19H8

A:Gene: CESP:F19H8.3
A:Map position: 2

A:Introns: 143/2

C:Superfamily: ADP-ribosylation factor

C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop

F.24-31/Region: nucleotide-binding motif A (P-loop)

F.90-95/Region: nucleotide-binding motif B

F.127-130/Region: GTP-binding NKXD motif

F.2/Modified site: myristylated amino end (gly) (in mature form) #status predicted

Query Match 5.6%; Score 75.5; DB 2; Length 184;
Best Local Similarity 21.5%; Pred. No. 49;

Matches 46; Conservative 40; Mismatches 79; Indels 49; Gaps 11;

OY 30 EILSFNREERFPMASTFFKVLGCAYLSRIDAGEOIGRIRHYSONDLVEYSPTEKHLT 89
DB 5 DYLSKFSK-----PGGREIRILLGL-----DNAGFTTILKQL--SSEVQVHTP-----T 48
OY 90 DGMTVRELCSNAITMSDNTAANLTLTTIGPKELTAFLHNMGDHV-----TRL 137
DB 49 KGFNVKTYAA-----MGD---IRLNVWDIGGQSTIRPYMSYENIDTLIFVIDSNDKRRF 101
OY 138 DRWEPELNEALPNDERTTTPVAMATTLRLKLTGELTLASROQLIDMWADRVAGPLR 197
DB 102 DENNIEIGELL-DEKLRKVPYLVIFANKQ-----DLVTAASSEITRKMLD-----LIR 150
OY 198 SALPAGFIADKSGAGERSGRIIALGPDGKRS 231
DB 151 DRT---WHIQACSAALKNEGINDIGTWAVASMLKPA 181

RESULT 5

T36975

hypothetical protein SCJ11.04 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36975

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1999

A:Reference number: Z21618

A:Accession: T36975

A:Status: preliminary; translated from GB/EMBL/DDJ

A:Molecule type: DNA

A:Residues: 1-195 <DLI>

A:Cross-references: EMBL:AL109949; PIDN:CAB52889.1; GSPDB:GN00070; SCOEDB:SCJ11.04

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCJ11.04

Query Match 5.5%; Score 74.5; DB 2; Length 195;
Best Local Similarity 17.0%; Pred. No. 64;

Matches 40; Conservative 43; Mismatches 77; Indels 75; Gaps 9;

OY 15 IGARVGYIEL-DLNGEILLESFRSEERFPMASTFFKVLGCAYLSRIDAGEOIGRIRHY 73
DB 17 MASMLGLRERREASAEVEVLR-----AARVAALAEAGEIELDRV--- 60
OY 74 QNDLVEYSPTEKHLTDGMTVRELCSNAITMSDNTAANLTLTTIGPRELTAFLHNMGDH 133
DB 61 -----IARELVEALAVSAETTYTEAGEGELTALVPAPASAEPPGI----- 104
OY 134 VTRLDWEPELEAL--PDNERTTTPVAMATTLRLKLTGELTLASROQLIDMWADRV 191
DB 105 -----VPHMOGLSVSVLSPNNOR-----ILNVLDQRPGLPEVPRANDI 142
OY 192 AGPLRSALPA-----GWFIDKSGAGERSGRIIALGPDGKRS 231
DB 143 AAALGTEAAAAAAGVEGRPRGKRLAERGLDPAASGAFSAGRR-LVAS--PGDPS 194

RESULT 6

AB3753

hypothetical protein BH0825 [Imported] - Bacillus halodurans (strain C-125)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 10.2857 Seconds
(Without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_37S

Perfect score: 1345
Sequence: 1 HPERLVKVKDAEDQAGARVG.....TMDERNQIAIGASLIRHW 263

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: PIR_76: *
2: PIR_1: *
3: PIR_3: *
4: PIR_4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	39.0	105	2 JC2566	b1a protein - Pseu
2	78	5.8	191	2 S67447	hypothetical prote
3	76.5	5.7	200	2 G97064	spore coat protein
4	75.5	5.6	184	2 T21126	ADP-ribosylation f
5	74.5	5.5	195	2 T36975	hypothetical prote
6	74	5.5	177	2 A83753	hypothetical prote
7	72.5	5.4	113	2 T45195	hypothetical prote
8	71.5	5.3	152	2 T36984	hypothetical prote
9	71	5.3	192	2 G83096	conserved hypotet
10	70.5	5.2	131	2 AD2281	hypothetical prote
11	70.5	5.2	180	2 C71869	hypothetical prote
12	70	5.2	108	2 E84217	hypothetical prote
13	69.5	5.2	145	2 F84251	hypothetical prote
14	69.5	5.2	167	2 D87360	conserved hypotet
15	68.5	5.1	172	2 AD3606	molybdopterin bios
16	68.5	5.1	190	2 T09136	ADP-ribosylation f
17	68	5.1	170	2 AB0192	conserved hypotet
18	68	5.1	172	2 T48420	hypothetical prote
19	67.5	5.0	192	2 A83587	hypothetical prote
20	67.5	5.0	198	2 D95285	conserved hypotet
21	66.5	4.9	128	2 F87353	conserved hypotet
22	66.5	4.9	188	1 G69365	conserved hypotet
23	66.5	4.9	196	2 H69687	hypothetical prote
24	66	4.9	146	2 C72703	nucleoside-diphosp
25	66	4.9	149	2 F71252	collagen alpha 1(X
26	65.5	4.9	442	2 F56978	ribosome binding f
27	65.5	4.9	116	2 G82906	probable bacteriof
28	65.5	4.9	195	2 AE0623	conserved hypotet
29	65.5	4.9	197	2 H90211	conserved hypotet

30	65	4.8	148	2 E75283	conserved hypotet
31	65	4.8	150	2 AC0284	probable heat shoc
32	65	4.8	153	2 E81708	conserved hypotet
33	65	4.8	160	1 E69186	conserved hypotet
34	65	4.8	168	2 B75498	conserved hypotet
35	65	4.8	180	2 G70912	hypothetical prote
36	65	4.8	198	2 C69256	conserved hypotet
37	64.5	4.8	151	2 D81333	probable hypotet
38	64.5	4.8	177	2 D90227	hypothetical prote
39	64	4.8	113	2 D70580	hypothetical prote
40	64	4.8	150	2 T08585	calmodulin - soybe
41	64	4.8	162	2 AG0769	probable acetyltra
42	64	4.8	168	2 C85715	unknown protein en
43	64	4.8	197	2 AC2133	iron(III) diclrat
44	63.5	4.7	168	2 T20606	hypothetical prote
45	63.5	4.7	179	2 AB1994	hypothetical prote

ALIGNMENTS

RESULT 1

JC2566 b1a protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweitzer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 14811281, 81-86, 1994

A>Title: Construction of improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A>Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MES>

C:Genetics:

A:Gene: b1a

C:Superfamily: beta-lactamase I

Query Match

Best local Similarity 100.0%; Pred. No. 9.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0;

OY	160	AAATTIRKLLTSELTLASROOLIDMENDKRVAGPLRSALPAGFIDKSGAGGRGSG	219
DB	2	AAATTIRKLLTSELTLASROOLIDMENDKRVAGPLRSALPAGFIDKSGAGGRGSG	61
OY	220	ITAAAGPDGKPSRIIVYITTSQATMDERNROIAETIGASLIRHW	263
DB	62	ITAAAGPDGKPSRIIVYITTSQATMDERNROIAETIGASLIRHW	105

RESULT 2

S67447 hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, March 1996

A:Reference number: 221766

A:Accession: T38062

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:Z69944; NID:q1217974; PIDN:CA93808.1; PID:q1217978; GSPDB:G

A:Experimental source: strain 972H-; cosmid c1F12

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match

5.8%; Score 78; DB 2; Length 191;

```

RESULT 15
US-10-106-698-4858
: Sequence 4858, Application US/10106698
: Publication NO. US20030109690A1
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
: FILE REFERENCE: PA00561
: CURRENT APPLICATION NUMBER: US/10/106,698
: CURRENT FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCR/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO 4858
: LENGTH: 179
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-106-698-4858

```

Query Match	4.8%	Score 65	DB 15	Length 179
Best Local Similarity	27.3%	Pred. No. 1,2e+02		
Matches	24	Conservative 18	Mismatches 34	Indels 12
			Gaps 4	
Qy	120	PKEITAFIHHNGDHYTRLD	-----RWPELNEA-----	INDDEKDTTPYPMAMATTLKRLLTG 171
Db	52	PRELTV-----GGEKIEYLDHSHKRMKLVANNEGRSGYISNIIIEPIQPEPTGTGGSPSRV 108		
Qy	172	ELIETLASH-QQLIDMEADKYAGPLLRS 198		
Db	109	PMRLRSSPEEVTDMIAENESTATVTR 136		

Search completed: September 10, 2003, 12:33:19
Job time : 18.5714 secs

Best Local Similarity 20.9%; Pred. No. 85;
Matches 39; Conservative 26; Mismatches 68; Indels 54; Gaps 7;
QY 50 ILGCAVLSRIDAGQGLGRIRHYSOND--LVEYSPYTEKHL-----TDCMTREL----- 97
Db 17 LLLAALVLTQTQADVPVPRATRLPEAKDCHIAQFKLSLPRELAFAKKAIAIKRLLEKDM 76
QY 98 -CSA-----AATMS--DNTANLLTTTGGPKELTAFLN 129
Db 77 RCSSHLISRAMDKOLQVGRPKALQAEVALTLKWMENINDSALTTTIGOLPHLTSHHS 136
QY 130 MGDHVT-----RLDRPELNEAIPNDRDTPVAMATTIRKLLTGLTL 176
Db 137 QLQTQLOATAEPKPPSRRLSRMRLQEA-OSKETPCLEDVTSNLFOLLRLDKCV 195
QY 177 ASRQOLI 183
Db 196 ASGDQCV 202

RESULT 12

US-09-764-870-390
; Sequence 390, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-390

Query Match 4.9%; Score 65.5; DB 9; Length 191;

Best Local Similarity 19.7%; Pred. No. 1.1e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;

QY 21 YIELDNGSEILSFSEERFPMMSYFKVLLCGAVLSRIDAGEOGLRIRHYSONDLVEY 80
Db 48 YCRSLNDSFV-----DMFGDYDSTFNSFAQVDDEOK-----YMQ----- 85
QY 81 SPYTEKHLTDGTVRELCASAITMSDNTANLLTTTGGPKELTAFLN----- 128
Db 86 LPEHKKATDFAT-ENLCSESI-----KNKLSITFTIGNLTELTQTKHTENOSGEGVTI 138
QY 129 -----NMGDHTRLDRMEPE-----LNEAIPND 151
Db 139 EPGADLLYDVSSQAIYFXNMLQNSSNDLGDHSKMDRWKSSSHNTVNEELPHN 191

RESULT 13

US-10-125-540-390
; Sequence 390, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-125-540-390

Query Match 4.9%; Score 65.5; DB 15; Length 191;

Best Local Similarity 19.7%; Pred. No. 1.1e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;

QY 21 YIELDNGSEILSFSEERFPMMSYFKVLLCGAVLSRIDAGEOGLRIRHYSONDLVEY 80
Db 48 YCRSLNDSFV-----DMFGDYDSTFNSFAQVDDEOK-----YMQ----- 85
QY 81 SPYTEKHLTDGTVRELCASAITMSDNTANLLTTTGGPKELTAFLN----- 128
Db 86 LPEHKKATDFAT-ENLCSESI-----KNKLSITFTIGNLTELTQTKHTENOSGEGVTI 138
QY 129 -----NMGDHTRLDRMEPE-----LNEAIPND 151
Db 139 EPGADLLYDVSSQAIYFXNMLQNSSNDLGDHSKMDRWKSSSHNTVNEELPHN 191

RESULT 14

US-09-764-868-757
; Sequence 757, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 757
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-757

Query Match 4.8%; Score 65; DB 10; Length 179;

Best Local Similarity 27.3%; Pred. No. 1.2e+02;
Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

QY 120 PKELTAFLNMGDHTRLD---RMEPELNEA-----IPNDRDTPVAMATTIRKLLTG 171
Db 52 PRELTVY---QGKLEVLDSKRMWLYKNAGSGSYIPSLLEPLQPGRTGCGQSPSRV 108
QY 172 ELITLASR-QQLIDMEADRVAGPLLR 198
Db 109 PMLRLSRPREVYDWLQAEENFSTATVRT 136

```

; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-765

Query Match
Best Local Similarity 24.7%; Score 67.5; DB 15; Length 184;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;

QY 41 FPMWSTKRV-----LCGAVLSRIDAGOEQLGRRIHXSONDLYEYSPV---TEKHLLTDG 91
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 23 FCMSTERRLLVYPYMLNGSVASRL-----NDSINGKRALDMPTRKRISLG 67
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 92 -----MTVRELCSAATMSDNTAANLL-----TTIGPKRELTAFLHMGDHTRLDRW-- 140
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 68 AARGLLVLEHOCCEKIIHRVYKANILDEYFEAVVGDGFLAKLLDRHDSHTVAVRGTV 127
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 141 ---EPELNEAIPNDERDTPVAVAMATLRLKLTGEL---LTLASROO--LIDMM 106
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 128 GHAPRELVSIGQSSER--TDVFGGIIILLITGQKALDGERAANKGVKMDWV 179
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 9
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: The111 et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948, 018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-018-8

Query Match
Best Local Similarity 5.0%; Score 67.5; DB 10; Length 187;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

QY 112 LLTTTGGPKELTAFLHMGDHTRLDRWEPE-----INEALPNDERDTPVAVAMAT- 163
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 11 LLLIALAPPEAS-----QYCGRIEYVWPNPNKCCSSCLORFEP-----PPCGALFETG 58
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 164 -----TLRKLGLGELLTLASR--QQLIDMMEA-----DKVAGP-----L 195
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 59 DTKREKSLPLSLSELSASQPLSRLLDELEVLEELLVLLDPGPGGGGNAHGTTRELA 118
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 196 LRSALPAGWFIADKSGAGERSGRIIALGPDGKPSR--IVVIYTTGSOATMDERNROIA 253
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 119 ARKGLPAAW-----STFAVSLRPSRSLPALILEMVAARPSASIGLGLTHIA 165
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 254 EIG 256
    :|::|
DB 166 QLG 168
```

```

RESULT 10
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 194;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

QY 202 AGWFIADKSGAGERSGSG-----ITAAVGPDG--RPSRIYVITYT 239
    :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 75 SGFVTRSGAGERYTEGAEYRITAAIGPLAVHEVRVAVVVT 118
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 11
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127, 816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 202;
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds
(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E_37S

Perfect score: 1345
Sequence: 1 HPELTLYKKVKAEDQDQARVCG.....TMDENRNOIAETGASLIKHV 263Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	463	34.4	94	US-10-102-806-760	Sequence 760, App
2	78	5.8	202	US-10-127-816-9	Sequence 9, App1
3	78	5.8	202	US-10-142-717-12	Sequence 12, App1
4	69.5	5.2	149	US-10-233-926-4	Sequence 4, App1
5	69	5.1	149	US-10-156-761-8136	Sequence 8136, App
6	69	5.1	178	US-09-738-626-5680	Sequence 5680, App
7	69	5.1	206	US-10-156-761-13867	Sequence 13867, A
8	67.5	5.0	184	US-10-101-464A-765	Sequence 765, App
9	67.5	5.0	187	US-09-948-018-8	Sequence 8, App1
10	67	5.0	194	US-10-156-761-8623	Sequence 8623, App
11	67	5.0	202	US-10-127-816-11	Sequence 11, App1
12	65.5	4.9	191	US-09-764-870-390	Sequence 390, App
13	65.5	4.9	191	US-10-125-540-390	Sequence 390, App
14	65	4.8	179	US-09-764-868-757	Sequence 757, App
15	65	4.8	179	US-10-106-698-4858	Sequence 4858, App

16	65	4.8	202	US-10-189-346-12	Sequence 12, App1
17	64.5	4.8	189	US-09-815-242-5862	Sequence 5862, App
18	64.5	4.8	189	US-09-815-242-12979	Sequence 12979, A
19	64.5	4.8	191	US-10-156-761-8978	Sequence 8978, App
20	64.5	4.8	193	US-09-805-354-8	Sequence 8, App1
21	64.5	4.8	193	US-10-144-259-8	Sequence 8, App1
22	63.5	4.7	202	US-10-156-761-13460	Sequence 13460, A
23	63.5	4.7	206	US-09-738-626-5425	Sequence 5425, App
24	62.5	4.6	144	US-10-131-406-4	Sequence 4, App1
25	62.5	4.6	160	US-09-882-227-230	Sequence 230, App
26	62.5	4.6	190	US-10-156-761-9507	Sequence 9507, App
27	62	4.6	162	US-09-738-626-4796	Sequence 4796, App
28	62	4.6	174	US-09-864-761-35777	Sequence 35777, A
29	62	4.6	185	US-10-101-464A-616	Sequence 616, App
30	62	4.6	190	US-09-738-626-4637	Sequence 4637, App
31	61.5	4.6	88	US-10-100-252-6	Sequence 6, App1
32	61.5	4.6	142	US-10-156-761-12111	Sequence 31, App1
33	61.5	4.6	187	US-10-189-346-16	Sequence 16, App
34	61.5	4.6	196	US-10-189-346-16	Sequence 16, App
35	61.5	4.6	202	US-10-189-346-16	Sequence 16, App
36	61	4.5	177	US-09-864-761-41555	Sequence 41555, A
37	61	4.5	177	US-09-791-932-69	Sequence 69, App
38	61	4.5	189	US-09-764-870-523	Sequence 523, App
39	61	4.5	189	US-10-125-540-523	Sequence 523, App
40	61	4.5	203	US-09-800-729-154	Sequence 154, App
41	60.5	4.5	152	US-10-156-761-14346	Sequence 14346, A
42	60.5	4.5	196	US-10-219-220-265	Sequence 265, App
43	60	4.5	134	US-09-768-2358-40	Sequence 40, App1
44	60	4.5	184	US-09-798-029-4	Sequence 4, App1
45	60	4.5	184	US-09-795-926-46	Sequence 46, App1

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103PIC1
; CURRENT APPLICATION NUMBER: US/10/102, 806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925, 298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-760

Query Match 34.4%; Score 463; DB 15; Length 94;
Best Local Similarity 98.9%; Pred. No. 2, 9e-40;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 75 NDIVEYSPYTEKHLLTDGTVRELCSSAITSNDTANLLTTGGPRELTAFLHNGDHY 134
DB 1 NDIVEYSPYTEKHLLTDGTVRELCSSAITSNDTANLLTTGGPRELTAFLHNGDHY 60

NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 24921
 LENGTH: 203
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24921

Query Match 4.88; Score 64.5; DB 4; Length 203;
 Best Local Similarity 27.18; Pred. No. 49;
 Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7;

QY 135 TRLDREPEL-----NEAIPNDERDTPVAMATILRLKLTGELLTASROQLDME 187
 DB 34 TPLDSSSRRLRVSLARAKAPRMAQARMPKLAET--RALTLSTLSSR---DMS- 86
 QY 188 ADVAGPLLSALPAGFIADKSGAGSGRGIITALGPDKPSPRIVIYTTGSQATWE 247
 DB 87 -----TSAMPGGDI-----SPGWLASIRLEBK-AKLAITATY---AALKV 123
 QY 248 RNRQIAETG 256
 DB 124 PTRRSTIG 132

RESULT 14

US-09-239-909-2
 Sequence 2, Application US/09239909
 Patent No. 6284952
 GENERAL INFORMATION:
 APPLICANT: Kumbo Petrochemical Co. Ltd.
 TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
 FILE REFERENCE: P99-2-6
 CURRENT FILING DATE: 1999-01-29
 EARLIER APPLICATION NUMBER: EP 99300136.1
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: KOPATIN 1.0
 SEQ ID NO 2
 LENGTH: 150
 TYPE: PRT
 ORGANISM: G. max calmodulin4 (SCAM4)
 US-09-239-909-2

Query Match 4.88; Score 64; DB 3; Length 150;
 Best Local Similarity 27.98; Pred. No. 35;
 Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SQNDLVEYSPVTEKHLTDG---MTVRELCSAATIMSNDNTAANLLTTI-----GSPKE 122
 DB 6 SEQIYDFKEAFGLPKDGGCITVELATVIRSLDNPTEELQDMISEVDADGNGTIE 65
 QY 123 LTFALNMGDHYTRLDREPELNEAIPNDERDTPVAMATILRLKLT--TGELLTASRO 180
 DB 66 FDFELSLMAKVKYDTDA--EELKEAFKVFEDKONGYIS--ASELRHVWVINGEKLTDDEYE 123
 QY 181 QLIDRMEAD 189
 DB 124 QMT--KEAD 130

RESULT 15

US-08-624-677A-2
 Sequence 2, Application US/08624677A
 Patent No. 6476192
 GENERAL INFORMATION:
 APPLICANT: Lally, Nicola C.
 APPLICANT: Jenkins, Mark C.
 APPLICANT: Dudey, Jitender P.
 TITLE OF INVENTION: Antigens useful for the serodiagnosis of
 TITLE OF INVENTION: Neosporosis
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Graeter, Janelle S.
 STREET: Rm. 411, Bldg. 005, BARC-W
 CITY: Beltsville
 STATE: MD
 COUNTRY: USA
 ZIP: 20705

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/624,677A
 FILING DATE: 15-Apr-1996
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Graeter, Janelle S.
 REGISTRATION NUMBER: 35,024
 REFERENCE/DOCKET NUMBER: 0228.95
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-504-6629
 TELEFAX: 301-504-5060

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 203 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-624-677A-2

Query Match 4.88; Score 64; DB 4; Length 203;
 Best Local Similarity 23.88; Pred. No. 56;
 Matches 39; Conservative 23; Mismatches 72; Indels 30; Gaps 7;

QY 54 AVLSRIDAGEQOLGRRIHYSQNDLVEYSPVTEKHLTDGNTVRELCSAA--ITMSDNTAAN 111
 DB 33 AGVSNTDGGDDDAAGNPVD-----SDVTDAITDGEPRVVSQCKPHTTONGSLIRK 83
 QY 112 LLLTTIGGPRELAFALNMGDHY--TRLDREPELNEAIPNDER-DTTPVA----- 160
 DB 84 LAVPVGA---LTSYL--VADRYLPELTSABEGTESIPCKKVKTRAVGIALVAAAAVA 138
 QY 161 ---MATTILRLKLTGELLTASROQLIDWMEADRVAGPLRSALP 201
 DB 139 GLGLARTFRHFVPRKSKTVASSEDALGNSSEQYVEGTVNGSSDP 182

Search completed: September 10, 2003, 12:31:04
 Job time : 12.8571 secs

```

? TELECOMMUNICATION INFORMATION
? TELEPHONE: 212-867-0123
?
? TELEFAX: 212-878-9655
?
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS
? LENGTH: 174 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? JS-09-262-666-6

```

Query Match	4.8%	Score 65	DB 4	Length 174
Best Local Similarity	23.9%	Pred. NO. 34		
Matches 28	Conservative 22	Mismatches 51	Indels 16	Gaps 4

```

0Y      7  KVNKDDEDLGAR-VGYTELDNNGSGLTSPFSEEEFPMMSPFKVLTGCAVLSRIDAGEO 65
Db      67  KYEQAAETELKEKNIFLVKVDCTEEALNCRDQVSEYIFPLIKIFRGL-----DAVPRY 117
0Y      66  LGR-----IHS-ONDLVETSPYTEKHLTGCAITVRELCSAATMSDNTANILLTT 116
Db      118  QGARTTEAIVSMWQSLPAVSPSPENLEIKTKDKIVLIGYIASDDQTAIDITFT 174

```

RESULT 11
US-08-557-122A-12
; Sequence 12, Application US/08557122A

GENERAL INFORMATION: Hjort, Carsten Mallard
APPLICANT: Hjort, Carsten Mallard
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: NO. 5879664d NO. 5879664dsk of NO. 5879664th America, Inc
STREET: 405 Lexington Avenue, 64th Floor

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: Patentin Release #1.0, Version #1.30
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/557,122A
8  FILING DATE: 11-DEC-1995

```

```

;
;   HIGHLIGHT/ASCII INFORMATION.
;
;   NAME: Lambdlfb, Elias J.
;
;   REGISTRATION NUMBER: 33,728
;
;   REFERENCE/DOCKET NUMBER: 3980-204-US
;
;   TELECOMMUNICATION INFORMATION:
;
;       TELEPHONE: 212-867-0123
;
;       TELEFAX: 212-878-9655
;
;   INFORMATION FOR SEQ ID NO: 12:
;
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 200 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;
;   MOLECULE TYPE: protein
;
US-08-557-122A-12
```

Db 94 QGARQTEAIVSYMKQSLPAVSPVTPENLEEIKTMDKIVVIGYIASDDQTANDIFTT 150

RESULT 12
US-09-262-666-12
; Sequence 12, Application US/09262666
; Patent No. 6346244

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;

```

Query Match	4.88;	Score 65;	DB 4;	Length 200;
Best Local Similarity	23.98;	pred. NO. 42;		
Matches 28;	Conservative	51;	Indels 16;	Gaps 4.;

QY 7 KVKDDEDLGAR-AGVIEYILDNSEIIESPFSEERPFPMSTFVYLGCAGVLSRIDGQEG 65
Db 43 KYEQAAETLEKKNIPVAVDCTEEBALCRQGVGCTYTLIFRGL-----DAVKPR 99
QY 66 LGRR-----IHS-ONDIYVSPYTEKHLTDGMATVRELCSAATMSDNTAANTLLTT 116
Db 94 QGARGTEAIVSYMKQSLPAVSPYTPETLEIKIKMDIVIGYIASDQVANDIFETT 150

```

RESULT 13
US-09-252-991A-74921
; Sequence 24921, Application US/09252991A
; Patent No. 6531795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 10/196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

```

```

OY 15 GANVGYELDNLNGEILLESFRSEERFPMNSTFYKVLGAVLSRIDAQO--EOLGRRIHY 72
Db 1 LKRNIGYLOOLSLIRELLSPISPCTKKXMSRFK-----AAAAPVDADSFIOELQRND- 54
OY 73 SONDLYEVPVTEHLTDGMTVRETCSAAITMSDWTANLLTTIGGRKELTAFLHNNGD 132
Db 55 -----SPYSEKSSSESTGROLLAPARIVASO-----PKLII----- 86
OY 133 HTVTLDMWPEPLNAPFNDERDOTTMPYAMATTLRLLTGELLTTLASROOLIDMNEADREVA 192
Db 87 -----LDEATANINDESETSLV--QASLAKMRGR--TTIAIHRLESTIDQANCI- 131
OY 193 GPLLRSLAPGMFTADKSGAGERSGRIIATLG 225
Db 132 -----YLDGGRILLESCTHELLALVG 152

```

```

: RESULT 8
: US-09-252-991A-29942
: Sequence 29942, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 29942
: LENGTH: 178
: TYPE: PRY
: ORGANISM: Pseudomonas aeruginosa
: FEATURE:
: NAME/KEY: UNSURE
: LOCATION: (116)
: OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
: US-09-252-991A-29942

```

RESULT 9
US-08-557-122A-6
Sequence 6: Application US/08557122A
Patent No. 5879664
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Mallaad
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5879664dsk of No. 5879666th America, Inc

```

Query Match      4.88; Score 65; DB 2; Length 174;
Best Local Similarity 23.98; Pred. No. 34;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4

OY 7 KYADAEQDQGAR-VGYIEIDLNSGETLSPFRSEEPFMSPFXKLCAGATLSIDAGQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 67 KREQAATLAKENINPLVAVDCTEEBALCRDQGVGGYPTLILIFRGL-----DAVRPY 117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 66 LGR-----IHS-ONDIYEVSPYTERKHLTGATVRELSAATMSDNTAANLLTT 116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 118 OGCRQTEAIVSYMWKSLPAVSPYTPENLEIKIMDLIVITGIVASDDQTDANDIFTT 174
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 10
US-09-262-666-6
: Sequence 6, Application US/09262666
: Patent No. 6346244
: GENERAL INFORMATION:
: APPLICANT: Hjort, Carsten Mølland
: TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
: NUMBER OF SEQUENCES: 38
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/262,666
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/557,122
: FILING DATE: 11-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Lambdlis, Elias J.
: REGISTRATION NUMBER: 33,728
: REFERENCE/DOCKET NUMBER: 3980,204-US

Db 111 SMADIAVAVTWSVITFTALPAAHSHATATASSRC 147

RESULT 5
US-09-413-814-87

; Sequence 87, Application US/09413814
; Patent No. 6223064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoffe, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 87
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match 4.9%; Score 66.5; DB 3; Length 198;
Best Local Similarity 22.4%; Pred. No. 28;
Matches 62; Conservative 31; Mismatches 77; Indels 107; Gaps 15;

QY 1 HPETLVKVRDAEQLGARVGYIELDLNSGEILSFERSERFPMMSF-----KVLTCGAV 55
DB 7 NPAAVAVAVKDKTS--AAFGV-----DQTVKRVLEIGIAQYVERAIVTFGEHFSYCVAS 60
QY 56 LSRIDAGOEQLGRHRYSONDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTANLLIT 115
DB 61 LVR-----MGVRIIRAA---TDROAD--IIR 82
QY 116 TIGCPKELTAFLHNMGDHYTRLDWPEPELNEALPNDERTMVAAMATLRLKLLGELT 175
DB 83 AVG-----ATRVQLQLETEMGRRV---GADITMPLAQ-----DLID 114
QY 176 LASROOLIDMEADKVAFLRLSALPAGWFIADK---SGAGER-----GSRGIITALG 225
DB 115 LASHRYRVPW-----NAHGPLVGOTL-AGSIRQRHRYHNVGVAPHNKRGRDPRLEAPT 169
QY 226 PDGKPSRIIVYITGSOATM-----DERNRQIAEIG 256
DB 170 PD-----YVIRDDITLLVGDSDSDVSRYFAEYG 197

RESULT 6
US-09-010-809-19

; Sequence 19, Application US/09010809B
; Patent No. 6090601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Belach, Mary C.
; TITLE OF INVENTION: Epothilone polyketide Synthases and Encoding DNA
; FILE REFERENCE: 30062-20020-00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-010-809-19

Query Match 4.9%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 26;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

QY 106 DNTANILLT-----TIGGKELTAPLHNMGDHYT--RLDRWP-----ELNEALPNDERO 154
DB 21 NHDARKHLLTSROGASAPGADVLRSELEALGASVTLAACVADPRALKDLIDNP----- 75
QY 155 TTPPVAMATLRLKLLTGCELTLASROOLIDMEADKVAFLRLSA-----LPAGWF 205
DB 76 SAHVVAAVVAASVDDDLGASMLEKI-----DRVFAKIDAAHMLHQLTQDKPLAAT 129
QY 206 IADKSGAGGRCRG 219
DB 130 ILFSSVAGVLCSSG 143

RESULT 7
US-08-858-207A-519

; Sequence 519, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-May-1997
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Glimm, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090

; TELEX:
; INFORMATION FOR SEQ ID NO: 519:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: No. 6348328e
US-08-858-207A-519

Query Match 4.9%; Score 65.5; DB 4; Length 170;

```

STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202

Query Match      5.4%; Score 72.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 3;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8

QY      130 MGDHV--TRLDRIWEELN---EALPNDNRDT---TMPVAMATLRLKLTGELLTLASRQ 180
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      12 MGDDIGMRERKERWGTNTQCPLRVVPGGDESPILDGRASPEDDIIT--NLISPTIMSHPPS 69
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      181 QLIDIMEEA-DKVAGPLL-----RSALDPAGFIADKSGAGBSGSCTIAAIGPDGKRPR 232
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      70 RDDDWVEFPDALQCTAVVDATDATKATMPAVGCI-----CASTRGS-GILASLSPPROPAR 123
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 3
US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6114307
GENERAL INFORMATION:
APPLICANT: Jaspers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TREATMENT OF INFECTIONS: STIMULATING PANCREATIC ISLET CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:
```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E_37S

Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDERNRQIAETGASLIRKH 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTOUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Dackfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	5.5	169	4 US-09-328-352-4491	Sequence 4491, App
2	72.5	5.4	124	4 US-08-311-731A-202	Sequence 202, App
3	71	5.3	159	3 US-08-991-890-4	Sequence 4, Appl
4	70.5	5.2	197	4 US-09-252-991A-24241	Sequence 24241, A
5	65.5	4.9	198	3 US-09-413-814-87	Sequence 87, Appl
6	65.5	4.9	158	3 US-09-010-809-19	Sequence 19, Appl
7	65.5	4.9	170	4 US-08-858-207A-519	Sequence 519, App
8	65.5	4.9	178	4 US-09-252-991A-29942	Sequence 29942, A
9	65	4.8	174	2 US-08-557-122A-6	Sequence 6, Appl
10	65	4.8	174	2 US-09-262-666-6	Sequence 6, Appl
11	65	4.8	200	2 US-08-557-122A-12	Sequence 12, Appl
12	65	4.8	200	2 US-09-262-666-12	Sequence 12, Appl
13	64.5	4.8	203	4 US-09-252-991A-24921	Sequence 24921, A
14	64.5	4.8	150	3 US-09-239-809-2	Sequence 2, Appl
15	64	4.8	203	4 US-08-624-677A-2	Sequence 2, Appl
16	62.5	4.6	144	2 US-08-225-480-4	Sequence 4, Appl
17	62.5	4.6	144	2 US-09-118-445-4	Sequence 4, Appl
18	62	4.6	132	4 US-09-252-991A-22681	Sequence 22681, A
19	62	4.6	204	4 US-09-252-991A-17611	Sequence 17611, A
20	61.5	4.6	142	4 US-09-345-473E-31	Sequence 31, Appl
21	61	4.5	199	2 US-09-252-991A-30363	Sequence 30363, A
22	60.5	4.5	167	2 US-08-690-849-2	Sequence 2, Appl
23	60.5	4.5	167	2 US-09-004-053-2	Sequence 2, Appl
24	60	4.5	144	4 US-09-252-991A-25578	Sequence 25578, A
25	60	4.5	158	4 US-09-107-532A-4218	Sequence 4218, App
26	60	4.5	181	2 US-08-482-142-195	Sequence 195, App
27	60	4.5	181	2 US-08-478-572-195	Sequence 195, App

28	60	4.5	181	3 US-08-484-296-195	Sequence 195, App
29	60	4.5	184	4 US-09-795-926-46	Sequence 46, Appl
30	59.5	4.4	103	4 US-09-732-210-1282	Sequence 1282, Ap
31	59.5	4.4	166	4 US-09-134-001C-5269	Sequence 5269, Ap
32	59.5	4.4	160	4 US-09-252-991A-24737	Sequence 24737, A
33	59.5	4.4	168	2 US-08-598-873-3	Sequence 3, Appl
34	59.5	4.4	168	2 US-08-605-430-3	Sequence 3, Appl
35	59.5	4.4	171	4 US-09-252-991A-22351	Sequence 22351, A
36	59	4.4	136	4 US-09-252-991A-30622	Sequence 30622, A
37	59	4.4	148	4 US-09-134-001C-3640	Sequence 3640, Ap
38	59	4.4	171	4 US-09-107-532A-3979	Sequence 3979, Ap
39	59	4.4	180	4 US-09-194-905-5	Sequence 5, Appl
40	59	4.4	189	4 US-08-671-548C-48	Sequence 48, Appl
41	58.5	4.3	102	3 US-08-479-744A-47	Sequence 47, Appl
42	58.5	4.3	102	3 US-08-280-757B-47	Sequence 47, Appl
43	58.5	4.3	174	4 US-09-072-596-281	Sequence 281, App
44	58	4.3	72	4 US-08-858-207A-412	Sequence 412, App
45	58	4.3	168	1 US-08-451-947-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-4491
Sequence 4491, Application US/09328352
Patent No. 6562958

GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

THIS PAGE BLANK (USPTO)

XX Sequence 158 AA:

Query Match 5.9%; Score 79.5; DB 22; Length 158;
 Best local Similarity 33.3%; Pred. No. 5.5;
 Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

QY 110 ANLLTTIGPKELTAFLENNMGDHYTRLDREPEELNEATPNDERDTMPVAMATTLR--K 167
 | : : | | | : | : | | | : | | | : | | | : | | | : | | | :
 Db 80 ARMATSTSGAPKYSASLTARGASPWTGPEPSTRPAI-----TTSAPLAFROTILSGR 134
 QY 168 LITGELLTLASRQQLIDW 185
 | : : | | | : | : | | | : | | | : | | | : | | | :
 Db 135 GRMGSVVLLATRIPILRW 152

Search completed: September 10, 2003, 12:24:52
 Job time : 33.7143 secs

Db 62 DTGEYSKETOR-----AMLS-----NIVLPPT-----PSQDTEVLNML 99
 Qy 132 DHTRTDRKEPELNE---ALPDERDTPVPMATTLRLKLTGBELLASPOOLIDME- 187
 Db 100 ERLEQIQ-----ELNETRLALIVINMPTIP-----TLKERQALIERIKE 139
 Qy 188 ---ADRVAGPLRLSAL 200
 Db 140 NMPSDRIT--LLESSL 153
 RESULT 15
 AAU45138
 ID AAU45138 standard; Protein; 158 AA.
 AC AAU45138;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 De Propionibacterium acnes immunogenic protein #6034.
 XX
 KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KM dermatological; osteopathic; neuroprotectant.
 OS
 XX
 Propionibacterium acnes.
 PN MO200181581-AA2.
 PN
 PD 01-NOV-2001.
 PD
 PF 20-APR-2001; 2001WO-US12865.
 PF
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 PR
 (CORI-) CORIXA CORP.
 PA
 PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI: 2001-616774/71.
 DR N-PSDB; AAS59525.
 XX
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 PT
 XX
 PS Example 1; SEQ ID NO 6333; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting the
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published.pct_sequences.

Db 1 TLVKKDAEDQIGARVGYIELDLNSGKILSFRRPFRPMSTFKVLLCGAVLSRIDAG 60
 Oy 64 EQLGRRI-----HYSQNDLV 78
 |||||
 |||||
 Db 61 EQLGRRIITLMTWLTSTHOSKSL 85

RESULT 12

AAF60626
 ID AAF60626 standard; Protein; 119 AA.

XX AAF60626;
 AC

DT 25-MAR-2003 (updated)
 DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

KM Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 KM fusion protein; beta-lactamase.

XX Synthetic.
 OS

Key Location/Qualifiers
 FH Protein 1..63
 FT /label= beta-lactamase

FT Misc-difference 64..66
 FT /label= adaptor

FT Protein 67..119
 FT /label= beta-urogastrone

PN DE3523634-A.

PD 09-JAN-1986.

PF 02-JUL-1985; 85DE-3523634.

PR 02-JUL-1984; 84JP-0137691.

PA (EART) EARTH CHEM CO LTD.

XX Ackl S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
 PI Matsushiro S;

DR WPI: 1986-015031/03.

DR N-PSDB; AAN60630.

XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 PT and transformed cells contg. It.

XX Disclousure; Page 55; 92pp; German.

CC The fusion protein is less easily degraded by proteases and so
 CC protects beta-urogastrone and beta-lactamase collects in the periplasm

CC of E.coli. It is therefore easy to collect and purify the product.

CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 CC stomach acid secretion and promotes cell growth, so is useful for

CC treating ulcers and wounds. Previously the product was obtd. only
 CC in small amts. from human urine.

CC See also AAN60628, and 30-32.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 119 AA;

Query Match 14.2%; Score 191; DB 7; Length 119;
 Best Local Similarity 95.0%; Pred. No. 9.1e-12;
 Matches 38; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 HPETLVKVKDAEDQIGARVGYIELDLNSGKILSFRRSEER 40
 |||||
 |||||
 Db 24 HPETLVKVKDAEDQIGARVGYIELDLNSGKILSFRRSEER 63

RESULT 13

AAV92783

ID AAV92783 standard; Protein; 51 AA.

XX AAV92783;
 AC

DT 29-AUG-2000 (first entry)

DE Vtgs-beta-lactamase fusion protein (partial).

KM Vtgs; vitellogenin; secretory signal sequence; gene expression;
 KM oestrogen receptor binding protein; systemic circulation;

KM beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Vtgs

FT Cleavage-site 15..16
 FT Protein 22..51
 FT /note= "beta-lactamase mature protein"

PN WO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

PA (UYSI-) UNIV SINGAPORE NAT.

PA (LAMT/) LAM T J.

PI Ding JL, Tan NS, Ho B;

DR WPI: 2000-365615/31.

DR N-PSDB; AAA28507.

PT Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

XX Example 6; Fig 14A; 73pp; English.

CC A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pBADVtgsbeta-lactamase was

CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an estrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric

XX host organism.

XX Sequence 51 AA;

Query Match 10.5%; Score 141; DB 21; Length 51;
 Best Local Similarity 96.6%; Pred. No. 4.3e-07;
 Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 PETLVKVKDAEDQIGARVGYIELDLNSGK 30
 |||||
 |||||
 Db 23 PETLVKVKDAEDQIGARVGYIELDLNSGK 51

DB 140 VVGTYG 145

RESULT 10
ID ABG27935 standard; Protein; 101 AA.
XX ABG27935;
AC
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #27926.
DE
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HXSE-) HXSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR N-PSDB: AAS92122.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS
XX Claim 20: SEQ ID No 58294; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. Abg00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 101 AA;

Query Match 27.7%; Score 372; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 1e-30;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 DTTTPVAMATTIRKLTGELLTASRQQLIDWMEADKVPAPLRSALPAGFTADKSGAG 213
DB 10 DTTTPVAMATTIRKLTGELLTASRQQLIDWMEADKVPAPLRSALPAGFTADKSGAG 69

OY 214 ERGSRGIIAALGPD 227
DB 70 ERGSRGIIAALGPD 83

RESULT 11
ID ABG27919 standard; Protein; 88 AA.
XX ABG27919;
AC
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #27910.
DE
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HXSE-) HXSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI: 2001-639362/73.
DR N-PSDB: AAS92106.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS
XX Claim 20: SEQ ID No 58278; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. Abg00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 88 AA;

Query Match 24.2%; Score 325; DB 22; Length 88;
Best Local Similarity 80.0%; Pred. No. 6.6e-26;
Matches 68; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

OY 4 TLVKVDAEHOIGARVGYIELDINSSEILSEFSEERFPMASTFKYLICGAVYSIDAQ 63
DB 4 TLVKVDAEHOIGARVGYIELDINSSEILSEFSEERFPMASTFKYLICGAVYSIDAQ 63

```

XX AAB59052;
AC
XX
XX 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
OS
XX WO200055173-A1.
PN
XX 21-SEP-2000.
PD
XX 08-MAR-2000; 2000KO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI; 2000-611515/58.
DR N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Clatm 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiant activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemia; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 94 AA;
SQ

```

Query Match 34.4%; Score 463; DB 21; Length 94;
Best Local Similarity 98.9%; Pred. No. 3.2e-10;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 75 NDLYERSPTVERKLTGCMYVRELCSAATMSDNTAANLLTTGGCKELTAFELHNGDHY 134
DB 1 NDLYERSPTVERKLTGCMYVRELCSAATMSDNTAANLLTTGGCKELTAFELHNGDHY 60
OY 135 TRLDRWPELNEAIPNDERDTPVAMATT 164
DB 61 TRLDRWPELNEAIPNDERDTPVAMATT 90

```

```

RESULT 9
AAP60628
ID AAP60628 standard; Protein; 159 AA.
XX
XX AAP60628;
AC
XX
XX 25-MAR-2003 (updated)
DE 17-JUN-1991 (first entry)
XX
XX Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX /label= beta-lactamase
XX /label= 97.100
XX /label= adaptor
XX Protein 101..159
XX /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART ) EARTH CHEM CO LTD.
PA
XX Aoki S, Ohgai H, Horioka A, Hiramatsu H, Koumoto S, Nishimura A;
XX Matsushiro S;
PI
XX WPI; 1986-015031/03.
DR N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX and transformed cells contg. it.
XX
XX Disclosure; Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E.coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obt. only
XX in small amts. from human urine.
XX See also AAN60628, and 30-32.
XX (Updated on 25-MAR-2003 to correct PA field.)
CC
XX Sequence 159 AA;
SQ

```

Query Match 29.4%; Score 396; DB 7; Length 159;
Best Local Similarity 67.5%; Pred. No. 6.3e-13;
Matches 85; Conservative 9; Mismatches 20; Indels 12; Gaps 2;

```

OY 1 HPETLVKVADEADQLGARVGYIELDINSGEILSFPSSEERFPMSTFKVLLCGAVLSRID 60
DB 24 HPETLVKVADEADQLGARVGYIELDINSGEILSFPSSEERFPMSTFKVLLCGAVLSRID 83
OY 61 AGQEQIGRRIRHYSQNDLYERS-----PYTEKHLTDGMYVRELCSAATMSDNTAANL 112
DB 84 AGQEQIGRRIRHYSQNDLYERSAKRNSDSECPLESH---DGYCLHDGYCMYIEALDKYACNC 139
OY 113 LRTTIG 118
DB 113 LRTTIG 118

```



```

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228624.
PR 01-SEP-2000; 2000US-0228987.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0233080.
PR 12-SEP-2000; 2000US-0233081.
PR 14-SEP-2000; 2000US-0233168.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235844.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

```

```

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 03-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX PA (HUMA-) HUMAN GENOME SCT INC.
XX FI Rosen CA, Barash SC, Ruben SM;
XX DR WPI: 2001-465566/50.
XX DR N-PSDB; AAS41090.
XX PT Novel polypeptides and polynucleotides useful for diagnosing,
XX PT preventing, treating neural, immune system, muscular, reproductive,
XX PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX PT diseases.
XX PS Claim 11: SEQ ID NO 1216; 1180pp: English.
XX XX

CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AUA2915-AUA23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 129 AA:

```

```

Query Match 34.9%; Score 469; DB 22; Length 129;
Best Local Similarity 95.8%; Pred. No. 1,2e-40;
Matches 92; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```


PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244647.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0244677.
PR 08-NOV-2000; 2000US-0244678.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
DR N-PSDB; AAS41091.
XX
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
XX
PS claim 11; SEQ ID NO 1217; 1180bp; English.
XX
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the

CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 146 AA:

Query Match 41.9%; Score 563; DB 22; Length 146;
Best Local Similarity 96.5%; Pred. No. 2, 3e-50;
Matches 111; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 HPEFLVKKVDAEDDQAGRVGYELDLNNGELLESPRSEPRFPMSFPVLLCGAVLSRD 60
DB 28 HPEFLVKKVDAEDDQAGRVGYELDLNNGELLESPRSEPRFPMSFPVLLCGAVLSRD 87
OY 61 AGQEDLGRRIHYSQNDLVEYSPVTEKHLLDGTVEELCSAATMSDNTANILLT 115
DB 88 AGQEDLGRRIHYSQNDLVEYSPVTEKHLLDGTVEELCSAATMSDNTANILLT 142

RESULT 3
AAM90870
ID AAM90870 standard; Protein; 146 AA.
XX
AC AAM90870;
XX
DT 07-NOV-2001 (first entry)
XX
XX Human immune/hematopoietic antigen SEQ ID NO:18463.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.

PA (EART) EARTH CHEM CO LTD.
XX
PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Kounoto S, Nishimura A;
PI Matsushiro S;
XX
DR WPI: 1986-015031/03.
N-PSDB; AAN60631.
XX
PT New gene for expression of beta-urogastrone - its deriva., plasmid(s)
and transfected cells contg. lt.
XX
PS Disclosure; Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collectes in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
XX
SQ Sequence 196 AA;
Query Match 43.0%; Score 579; DB 7; Length 196;
Best Local Similarity 97.4%; Pred. No. 7.7e-52;
Matches 114; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 HPELVKDAEDQIGARVGYIELDNSGELSEFSESEFPFPMSTFKVLLGAVLSRID 60
DB 24 HPELVKDAEDQIGARVGYIELDNSGELSEFSESEFPFPMSTFKVLLGAVLSRID 83
QY 61 AGGEQIGRRHHYSONDLVEYSPYTEKHLTDGMTVREICSAITMSDNTAANLLTTI 117
DB 84 AGGEQIGRRHHYSONDLVEYSPYTEKHLTDGMTVREICSAITMSDNTAANLLTTI 140
RESULT 2
AAU23221 standard; Protein: 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0229519.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 : Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E_37S

Perfect score: 1345
Sequence: 1 HPTLVKVKDAEDQIGARVG.....TMDERNQIAEIGASLIRHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.GeneSeq.19Jun03.*

1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	579	43.0	196 7 AAP60627	Beta-urogastrone -
2	563	41.9	146 22 AAU23221	Novel human enzyme
3	563	41.9	146 22 AAM90870	Human immune/haema
4	484	36.0	127 18 AAM20440	H. pylori cytoplas
5	469	34.9	129 22 AAU23220	Novel human enzyme
6	469	34.9	129 22 AAM90871	Human immune/haema
7	469	34.9	182 22 ABG27917	Novel human diagno
8	463	34.4	94 21 AAB59052	Breast and ovarian
9	396	29.4	159 7 AAP60628	Beta-urogastrone -

10	372	27.7	101 22 ABG27935	Novel human diagno
11	325	24.2	88 22 ABG27919	Novel human diagno
12	191	14.2	119 7 AAP60626	Beta-urogastrone -
13	141	10.5	51 21 AAU92783	Vggs-beta-lactama
14	81	6.0	200 16 AAM55524	H. pylori ORF 29ep
15	79.5	5.9	158 22 AAU45138	Protonlactarium
16	78	5.8	202 24 AAE32763	Human zcyto24 prot
17	78	5.8	202 24 AAO16283	Mouse IMX129840-2
18	71	5.3	159 19 AAM65170	N-terminally tagged
19	70	5.2	15 10 AAM98503	Sequence encoded b
20	70	5.2	202 22 AAG82686	S. epidermidis ope
21	69.5	5.2	170 22 AAU56330	Protonlactarium
22	69	5.1	134 21 AAG27220	Zea mays protein f
23	69	5.1	178 22 AAG91926	C glutamylum prote
24	69	5.1	180 21 AAG20206	Arabidopsis thalia
25	69	5.1	194 22 AAG81764	S. epidermidis ope
26	67.5	5.0	177 21 AAB41582	Human ORFX ORF1346
27	67.5	5.0	184 21 AAB25446	Human ORFX ORF1346
28	67.5	5.0	187 23 AAU97104	Human MK61 protein
29	67	5.0	101 22 AAU66278	Protonlactarium
30	67	5.0	168 22 ABE68717	Drosophila melanog
31	67	5.0	202 24 AAE32764	Human zcyto25 prot
32	66.5	4.9	196 21 AAU73344	Human zcyto25 prot
33	66	4.9	170 22 AAU52472	Protonlactarium
34	65.5	4.9	50 17 AAM03591	Human alpha 2 C4 a
35	65.5	4.9	102 23 AAB34479	Human dehydrogenas
36	65.5	4.9	136 21 AAU44229	Partial corn extra
37	65.5	4.9	152 22 ABG19138	Novel human diagno
38	65.5	4.9	152 22 ABG27771	Novel human diagno
39	65.5	4.9	170 19 AAM38722	S. pneumoniae SGHR
40	65.5	4.9	191 22 AAU87418	Human novel extrac
41	65.5	4.9	191 22 AAU19740	Human novel extrac
42	65.5	4.9	191 23 AAB47960	Human polypeptide
43	65	4.8	113 22 AAB49942	Protein disulfide
44	65	4.8	174 16 AAR59508	Aspergillus sp. tr
45	65	4.8	179 22 AAU87627	Novel central nerv

ALIGNMENTS

RESULT 1	AAU23221	196 AA.
AAU23221	AAU23221	standard; Protein: 196 AA.
XX	XX	
AC	AAU23221	
XX	XX	
DT	25-MAR-2003 (updated)	
DT	17-JUN-1991 (first entry)	
XX	XX	
DE	Beta-urogastrone - beta-lactamase fusion protein from pUG2101.	
DE	XX	
KW	Beta-urogastrone gene: hormone; salivary gland; ulcers; wounds;	
KW	fusion protein: beta-lactamase.	
OS	Synthetic.	
XX	XX	
FT	Key	Location/Qualifiers
FT	Protein	1..120
FT	Misc-difference	/label= beta-lactamase
FT	Protein	121..123
FT	Protein	/label= adaptor
FT	Protein	124..196
FT	Protein	/label= beta-urogastrone
PN	DE5323634-A.	
XX	XX	
PD	09-JAN-1986.	
XX	XX	
PF	02-JUL-1985;	85DE-3523634.
XX	XX	
PR	02-JUL-1984;	84UP-0137691.
XX	XX	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_157T

Sequence: 1 HPEITLVKVKDAEDQAGARVC.....TMDERNKQIAETGASLIRHW 263

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

1: SPTREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	38.9	105	2	052026
2	524	38.9	105	2	052330
3	499	37.0	145	2	08RTD8
4	488	36.2	102	2	052639
5	464	34.4	138	2	0990F3
6	464	34.4	138	2	0990F3
7	464	34.4	139	2	09AMA1
8	462	34.3	138	2	09AMA2
9	462	34.3	138	2	09AMA0
10	462	34.3	139	2	09AM98
11	459	34.1	139	2	09AMA3
12	338	25.1	67	2	053553
13	326	24.2	95	2	08CDE5
14	310	23.0	62	2	08CDE5
15	301	22.3	128	2	09UN58
16	251	18.6	180	2	08KVT2

17	250	18.5	52	2	09R412	09R412 shigella fl
18	241	17.9	48	2	09RLH0	09RLH0 proteus mtr
19	197	14.6	38	2	P97145	P97145 escherichia
20	168.5	12.5	134	2	08VTL3	08VTL3 staphylococ
21	123.5	9.2	109	2	053698	053698 staphylococ
22	105	7.8	20	2	P97146	P97146 escherichia
23	87	6.5	100	2	093505	093505 staphylococ
24	82	6.1	68	2	09XBJ2	09XBJ2 bacillus ce
25	80	5.9	198	2	09ACM8	09ACM8 streptococ
26	79.5	5.9	202	16	098520	098520 rhizobium l
27	77	5.7	153	5	08S520	08S520 clona intest
28	75	5.6	181	16	08D108	08D108 yersinia pe
29	74.5	5.5	113	2	049970	049970 mycobacteri
30	74.5	5.5	172	16	08BWM3	08BWM3 brucella me
31	74.5	5.5	172	16	08BWM3	08BWM3 brucella su
32	73	5.4	131	2	09X9H0	09X9H0 yersinia en
33	73	5.4	175	17	08RTH8	08RTH8 pyrobaculum
34	72.5	5.4	145	2	005984	005984 staphylococ
35	72.5	5.4	204	11	091YW4	091YW4 mus musculu
36	72	5.3	205	16	09BJ18	09BJ18 rhizobium l
37	71.5	5.3	152	16	09RI91	09RI91 streptomyc
38	71.5	5.3	190	5	076227	076227 trypanosoma
39	71.5	5.3	195	16	09RI10	09RI10 streptomyc
40	71.5	5.3	196	2	09AH34	09AH34 pseudomonas
41	71.5	5.3	200	16	097JE8	097JE8 clostridium
42	71	5.3	150	10	0941G4	0941G4 nicotiana t
43	71	5.3	153	16	09PKT4	09PKT4 chlamydia m
44	71	5.3	177	16	09KEM7	09KEM7 bacillus ba
45	71	5.3	192	16	09HM07	09HM07 pseudomonas

ALIGNMENTS

RESULT 1

ID	052026	PRELIMINARY;	PRT;	105 AA.
AC	052026;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Beta-lactamase (Fragment).			
GN	BLA.			
OS	Pseudomonas aeruginosa.			
OG	Plasmid pR01614.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;			
OC	Pseudomonadaceae; Pseudomonas.			
OX	NCBI_TaxID=287;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-95011664; PubMed-7926843;			
RA	West S.E., Schweizer H.P., Dall C., Runyen-Janecky L.J.;			
RT	*Construction of improved Escherichia-Pseudomonas shuttle vectors			
RT	derived from pUC18/19 and sequence of the region required for their			
RT	replication in Pseudomonas aeruginosa.;			
RL	Gene 148:81-86(1994).			
DR	EMBL; L30112; AAA66058.1; -			
DR	HSSP; P00810; 1XPB.			
DR	InterPro: IPR001466; Beta_lactamase.			
DR	InterPro: IPR000871; Beta_lactamase_A.			
DR	Pfam: PF00144; beta_lactamase; 1.			
DR	PRINTS; PR00118; BLACTAMASEA.			
KW	Plasmid.			
FT	NON_TER			
SO	SEQUENCE	1	1	
	105 AA;	11229 MW;	D2889A407330557 CRC64;	

Query Match 38.9%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 8.5e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATILRLKLTGELTLASROOLIDMMEADRVAGSLRSALPAGCFINDKSGAGRGSRG 219
DB 2 AMATILRLKLTGELTLASROOLIDMMEADRVAGSLRSALPAGCFINDKSGAGRGSRG 61

QY 220 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 263
 Db 62 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 105

RESULT 2

Q52330 PRELIMINARY; PRT; 105 AA.

ID 052330
 AC 052330
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Tni bta protein (Fragment).
 OS Escherichia coli.
 OC plasmid RK2.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90264294; PubMed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.:
 RT "The Kil-kor regulon of broad host-range Plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korc.";
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL; M32794; AAA26408.1; -.
 DR HSSP; P00810; 1XPB
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS; PR00118; BLACTMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;
 Query Match 38.9%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 8.5e-37;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 AMATTIRKLTLGELLTLASRQQLIDWMEADKVAGPLRLSALPAGWFIADKSGAGERSRG 219
 Db 2 AMATTIRKLTLGELLTLASRQQLIDWMEADKVAGPLRLSALPAGWFIADKSGAGERSRG 61
 QY 220 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 263
 Db 62 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 105

RESULT 3

Q8RTD8 PRELIMINARY; PRT; 145 AA.

ID 08RTD8
 AC 08RTD8
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Palisudramaniam S.:
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftazidime-resistant Klebsiella pneumoniae."
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER
 SQ SEQUENCE 145 AA; 145

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 37.0%; Score 499; DB 2; Length 145;
 Best Local Similarity 68.8%; Pred. No. 1.8e-34;
 Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 100 AATMSDNTANLTLTIGGPKELTAFLHNGDHYRLDWEPELNEALPNDERDITTPV 159
 Db 7 AAYMSDMSAANLLLATVGGPAGLTAFLRQIGDNTFLDRWETELNEALPGDARDITTPA 66
 QY 160 AMATTIRKLTLGELLTLASRQQLIDWMEADKVAGPLRLSALPAGWFIADKSGAGERSRG 219
 Db 67 SMATTIRKLTLTSQLRSARSQLQWVDDRVAGPLRLSVLPAGWFIADKSGAKRGAR 136
 QY 220 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 263
 Db 127 IVALIGNNKAERIVVITTTGSOATMDERNQIAEIGASLIIKH 144

RESULT 4

Q52639 PRELIMINARY; PRT; 102 AA.

ID 052639
 AC 052639
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Beta-Lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC plasmid PRO1600.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX TRANSPOSOM-T01;
 RX MEDLINE=9436757; PubMed=8058819;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.:
 RT "Deletion and transposon mutagenesis and sequence analysis of the pOR
 RT PRO1600 OriR region found in the broad-host-range plasmids of the pOR
 RT series."
 RL Plasmid 31:265-274(1994).
 DR EMBL; L22691; AAA98312.1; -.
 DR HSSP; P00810; 1XPB.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS; PR00118; BLACTMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.2%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 9e-34;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTIRKLTLGELLTLASRQQLIDWMEADKVAGPLRLSALPAGWFIADKSGAGERSRG 219
 Db 2 AMATTIRKLTLGELLTLASRQQLIDWMEADKVAGPLRLSALPAGWFIADKSGAGERSRG 61
 QY 220 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 263
 Db 62 IIALGPDGKPSRIVVITTTGSOATMDERNQIAEIGASLIIKH 99

RESULT 5

Q990F3 PRELIMINARY; PRT; 138 AA.

ID 0990F3
 AC 0990F3
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)


```

DE SHV beta-lactamase (Fragment).
OC Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327049; AK07464.1; -
DR EMBL: AF327051; AK07466.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

```

```

Query Match 34.4%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 1.5e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 111 NLLTTIGPKELTAFLHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAMATTLRLLT 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDADDTTTPASMAATLRLLT 69
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 171 GELLTLASROQLIDMNEADKVNAPLLRSALPAGWFIADKSGGERSGCIITAAAGPDGKP 230
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGARGIVALLGPNNKA 129
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 231 SRIVVIY 237
   |||||
DB 130 ERIVVIY 136

```

RESULT 6

```

Q9AMA1 PRELIMINARY; PRT; 138 AA.
AC Q9AMA1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IB/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327050; AK07465.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

```

```

Query Match 34.4%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 1.5e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 111 NLLTTIGPKELTAFLHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAMATTLRLLT 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDADDTTTPASMAATLRLLT 69
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 171 GELLTLASROQLIDMNEADKVNAPLLRSALPAGWFIADKSGGERSGCIITAAAGPDGKP 230
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGARGIVALLGPNNKA 129
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 231 SRIVVIY 237
   |||||
DB 130 ERIVVIY 136

```

RESULT 7

```

Q9AMA2 PRELIMINARY; PRT; 139 AA.
AC Q9AMA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327048; AK07463.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

```

```

Query Match 34.4%; Score 464; DB 2; Length 139;
Best Local Similarity 68.5%; Pred. No. 1.5e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 111 NLLTTIGPKELTAFLHNGDHYTRLDRWEPELNEAIPNDERDTTTPVAMATTLRLLT 170
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 10 NLLATVGGPAGLTAFLRQIGDNYTRLDRWETELNEALPDADDTTTPASMAATLRLLT 69
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 171 GELLTLASROQLIDMNEADKVNAPLLRSALPAGWFIADKSGGERSGCIITAAAGPDGKP 230
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGARGIVALLGPNNKA 129
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 231 SRIVVIY 237
   |||||
DB 130 ERIVVIY 136

```

RESULT 8

```

Q9AMA0 PRELIMINARY; PRT; 138 AA.
AC Q9AMA0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-E/98 9-1;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from cefotaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327052; AK07467.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; Beta_lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 138 AA; 15305 MW; 561D092F5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 2.2e-31;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNMGDHVTRLDRMEPELNEAIPNDERDTTPVAMATTLRLKLLT 170
 |||||:||||| ||||| :||:||||||| |||||:| |||||:| |||||
 DB 10 NLLATVGGPAGLTAFLROIGDNTRLDRMETELNEALPGDARTTTPASMAATLRLKLLT 69
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 OY 171 GELLTLASROQLIDMEADKVGPLRLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 DB 70 SQRLSARSQROLQWNVDDRVAGPLIRSVLPAGWFIADKSGARGIALLGPNNKA 129
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 OY 231 SRIVVIY 237
 |||||:| |||||:| |||||:| |||||:| |||||:| |||||
 DB 130 ERIVVLY 136

RESULT 9

O9AM99 PRELIMINARY; PRT; 138 AA.

AC O9AM99.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/99 4-1;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from cefotaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327053; AK07468.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta_lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 138 AA; 15204 MW; 5609AC3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
 Best Local Similarity 67.7%; Pred. No. 2.2e-31;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNMGDHVTRLDRMEPELNEAIPNDERDTTPVAMATTLRLKLLT 170
 |||||:||||| ||||| :||:||||||| |||||:| |||||:| |||||
 DB 10 NLLATVGGPAGLTAFLROIGDNTRLDRMETELNEALPGDARTTTPASMAATLRLKLLT 69
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 OY 171 GELLTLASROQLIDMEADKVGPLRLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 DB 70 SQRLSARSQROLQWNVDDRVAGPLIRSVLPAGWFIADKSGARGIALLGPNNKA 129
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||

OY 231 SRIVVIY 237
 |||||:| |||||:| |||||:| |||||:| |||||:| |||||
 DB 130 ERIVVLY 136

RESULT 10

O9AM98 PRELIMINARY; PRT; 139 AA.

AC O9AM98.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/99 3-2;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from cefotaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327054; AK07469.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta_lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 139 AA; 15234 MW; 03361A792F5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
 Best Local Similarity 67.7%; Pred. No. 2.2e-31;
 Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNMGDHVTRLDRMEPELNEAIPNDERDTTPVAMATTLRLKLLT 170
 |||||:||||| ||||| :||:||||||| |||||:| |||||:| |||||
 DB 10 NLLATVGGPAGLTAFLROIGDNTRLDRMETELNEALPGDARTTTPASMAATLRLKLLT 69
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 OY 171 GELLTLASROQLIDMEADKVGPLRLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 DB 70 SQRLSARSQROLQWNVDDRVAGPLIRSVLPAGWFIADKSGARGIALLGPNNKA 129
 :||:||||| |||||:| |||||:| |||||:| |||||:| |||||
 OY 231 SRIVVIY 237
 |||||:| |||||:| |||||:| |||||:| |||||:| |||||
 DB 130 ERIVVLY 136

RESULT 11

O9AM93 PRELIMINARY; PRT; 139 AA.

AC O9AM93.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-935;
 RA Subramaniam G., Navaratnam P.;
 RT SHV gene sequences from cefotaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.*;
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327047; AK07462.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds

(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E

Perfect score: 1348
Sequence: 1 HPEITLVKVADEQICARVG.....TMDERNKQIAETIGASLIKHW 263

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 segs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	587	43.5	196	7	AAp60627
2	571	42.4	146	22	AAU23221
3	571	42.4	146	22	AAU23221
4	492	36.5	127	18	AAU20440
5	477	35.4	129	22	AAU23220
6	477	35.4	129	22	AAU23220
7	477	35.4	182	22	AAU23220
8	463	34.3	94	21	AAU23220
9	404	30.0	159	7	AAp60628

10	372	27.6	101	22	ABG27935	Novel human diago
11	333	24.7	88	72	ABG27919	Novel human diago
12	199	14.8	119	7	AAp60626	Beta-urogastrone -
13	141	10.5	51	21	AAU2783	Vtgs-beta-lactama
14	79.5	5.9	158	22	AAU45138	Propionibacterium
15	79.5	5.9	200	18	AAU55524	H. pylori ORF 29ep
16	78	5.8	15	10	AAU98503	Sequence encoded b
17	78	5.8	202	24	AAU32763	Human zcyto24 prot
18	78	5.8	202	24	AAU32763	Mouse IMX12940-2
19	73.5	5.5	170	19	AAU38122	S. pneumoniae SGHR
20	71	5.3	159	19	AAU69170	N-terminally tagge
21	70	5.2	202	22	AAU62866	S. epidermidis ope
22	69.5	5.2	170	22	AAU56330	Propionibacterium
23	69	5.1	134	21	AAU62720	zeamays protein f
24	69	5.1	178	21	AAU91925	C glutamicum prote
25	69	5.1	180	21	AAU20206	Arabidopsis thalia
26	69	5.1	194	22	AAU81764	S. epidermidis ope
27	67.5	5.0	184	21	AAU25446	Pinus radiata cell
28	67.5	5.0	187	21	AAU05495	Arabidopsis thalia
29	67.5	5.0	187	21	AAU05495	Arabidopsis thalia
30	67.5	5.0	187	23	AAU97104	Human MK61 protein
31	67	5.0	101	22	AAU66278	Propionibacterium
32	67	5.0	168	22	AAU66278	Drosophila melanog
33	67	5.0	202	24	AAU52472	Human zcyto25 prot
34	66	4.9	170	22	AAU52472	Propionibacterium
35	66	4.9	196	18	AAU14564	Streptococcus pneu
36	65.5	4.9	50	17	AAU03591	Human alpha 2 C4 a
37	65.5	4.9	102	23	AAU34479	Human dehydrogenas
38	65.5	4.9	136	21	AAU44429	Partial corn extra
39	65.5	4.9	152	22	AAU19138	Novel human diago
40	65.5	4.9	152	22	AAU27771	Novel human diago
41	65.5	4.9	177	21	AAU41582	Human ORF ORF1346
42	65.5	4.9	191	22	AAU87418	Novel central nerv
43	65.5	4.9	191	22	AAU19740	Human novel extrac
44	65.5	4.9	191	23	AAU47960	Human polypeptide
45	65.5	4.9	196	21	AAU73544	HRM clone 0258181

ALIGNMENTS

RESULT 1	AAp60627	standard; Protein: 196 AA.
ID	AAp60627	standard; Protein: 196 AA.
XX		
AC	AAp60627	
XX		
DT	25-MAR-2003	(updated)
DT	17-JUN-1991	(first entry)
XX		
DE	Beta-urogastrone - beta-lactamase fusion protein from pUG2101.	
XX		
KW	Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;	
KW	fusion protein; beta-lactamase.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Protein	1..120
FT	Misc-difference	/label- beta-lactamase
FT	Protein	121..123
FT		/label- adaptor
FT		124..196
FT		/label- beta-urogastrone
XX		
PN	DE3523634-A.	
XX		
XX	09-JAN-1986.	
PD		
XX	02-JUL-1985;	85DE-3523634.
PF		
XX	02-JUL-1984;	84JP-0137691.
PR		
XX		

PA (EART) EARTH CHEM CO LTD.
XX
PI Aoki S, Oheal H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX
DR WPI; 1986-015031/03.
DR N-PSDB; AAN60631.
XX
PT New gene for expression of beta-urogastrone - its derive, plasmid(s)
PT and transformed cells contg. It.
XX
PS Disclosure; Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collecta in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which supresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 196 AA:

Query Match 43.5%; Score 587; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. No. 3.3e-53;
Matches 115; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPELVYKDAEDDGLARVGYIELDNGSELSEFRRPRPMSTKVLGAVLSRID 60
DB 24 HPELVYKDAEDDGLARVGYIELDNGSKLSEFRRPRPMSTKVLGAVLSKVD 83

QY 61 AGQQLGRIRHYSONDLVEYSPYTEKHLDGMYRELCNAITMSDNTAALLTTI 117
DB 84 AGQQLGRIRHYSONDLVEYSPYTEKHLDGMYRELCNAITMSDNTAALLTTI 140

RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW lysase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
XX OS
XX WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225470.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR	14-AUG-2000	2000US-0225266
PR	14-AUG-2000	2000US-0225267
PR	14-AUG-2000	2000US-0225268
PR	14-AUG-2000	2000US-0225269
PR	14-AUG-2000	2000US-0225270
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	14-AUG-2000	2000US-0225799
PR	14-AUG-2000	2000US-0226129
PR	12-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226682
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-SEP-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231966
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232396
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232427
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	13-OCT-2000	2000US-0239660
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0244675
PR	08-NOV-2000	2000US-0244676
PR	08-NOV-2000	2000US-0244677
PR	08-NOV-2000	2000US-0244678
PR	08-NOV-2000	2000US-0245523
PR	08-NOV-2000	2000US-0245524

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

XX PA
XX FI Rosen CA, Barash SC, Ruden SM;
XX DR WPI; 2001-483426/52.
DR N-PADB; AAK63651.
XX XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
PS PS

XX Claim 11; SEQ ID NO 18463; 3071bp + Sequence Listing; English.
XX XX

CC AAk64951 to AAk64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAk82170 to AAk91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (II)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAk64703
CC to AAk67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAk54942 to AAk54950 and AAk62169
CC represent sequences used in the exemplification of the present invention
XX XX

SQ Sequence 146 AA;
 Query Match 42.4%; Score 571; DB 22; Length 146;
 Best Local Similarity 97.4%; Pred. No. 1e-51;
 Matches 112; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 HPETLVKVAEDQQLGARVGYIELDLSGKILESFRPERFPMMSTFKVLLCGAVLSRID 60
 DB 28 HPETLVKVAEDQQLGARVGYIELDLSGKILESFRPERFPMMSTFKVLLCGAVLSRID 87
 QY 61 AGQEQGLRRIRHYSONDLVEYSPYTERKHLTDGMTVRELCSAATMSDNTAANLLLT 115
 DB 88 AGQEQGLRRIRHYSONDLVEYSPYTERKHLTDGMTVRELCSAATMSDNTAANLLLT 142

RESULT 4
 AAM20440
 ID AAM20440 standard; protein; 127 AA.
 AC AAM20440;
 DT 14-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein, 34574062.aa.
 XX
 KM Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KM binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KM duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
 KM replication; transcription; recombination; repair.
 XX Helicobacter pylori.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 121 /note="encoded by codon YGC"
 FT Misc-difference 122 /note="encoded by codon WGC"
 XX
 PN MO9640893-A1.
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BU;
 XX
 DR WPI: 1997-052306/05.
 DR N-PSDB; AAT67613.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61: Pages 615-616; 1481pp; English.
 XX
 CC The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in genomic replication, transcription, recombination and repair.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed for ORF of at least 180 nucleotides, and the
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 SQ Sequence 127 AA;
 Query Match 36.5%; Score 492; DB 18; Length 127;
 Best Local Similarity 99.0%; Pred. No. 1.7e-43;
 Matches 96; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPETLVKVAEDQQLGARVGYIELDLSGKILESFRPERFPMMSTFKVLLCGAVLSRID 60
 DB 24 HPETLVKVAEDQQLGARVGYIELDLSGKILESFRPERFPMMSTFKVLLCGAVLSRID 83
 QY 61 AGQEQGLRRIRHYSONDLVEYSPYTERKHLTDGMTVREL 97
 DB 84 AGQEQGLRRIRHYSONDLVEYSPYTERKHLTDGMTVREL 120

RESULT 5
 AU23220
 ID AU23220 standard; protein; 129 AA.
 AC AU23220;
 DT 18-DEC-2001 (first entry)
 DE H. pylori cytoplasmic protein, 34574062.aa.
 XX
 KM Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KM ligase; hyperproliferative disorder; immunodeficiency disorder;
 KM autoimmune disorder; neurological disorder; metabolic disorder;
 KM inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KM blood-related disorder; infectious disorder; cytostatic; anti arthritic;
 KM neoplastic; anticoagulant.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc-difference 121 /note="encoded by codon YGC"
 FT Misc-difference 122 /note="encoded by codon WGC"
 XX
 PN WO200155301-A2.
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0196123.
 PR 19-MAY-2000; 2000US-020515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.

OY 1 HPELVKVKDAEDLGARVGIETLDNGETLESFREPERRPMSTFKVLGAVLSRID 60
DB 28 HPELVKVKDAEDLGARVGIETLDNSCKILLESFREPERRPMSTFKVLGAVLSRID 87
OY 61 AGQEOLGRRIHYSNDLVEYSPTVEKHILTOGMTATRE 96
DB 88 AGQEOLGRRIHYSNDLVEISPTVEKHILTOGMTATRE 123

RESULT 6
AAM90871
ID AAM90871 standard; Protein: 129 AA.
XX
AC AAM90871;
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18464.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-0501354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.


```

XX AAB59052;
XX
XX 27-MAR-2001 (first entry)
XX
XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
XX Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neutropenic; neurprotection; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antineoplastic; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX W02005173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05861.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-611515/58.
XX
XX N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antineoplastic; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiac activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemia; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 94 AA;
XX
XX Query Match 34.3%; Score 463; DB 21; Length 94;
XX Best Local Similarity 98.9%; Pred. No. 1,2e-40;
XX Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 75 NDLYEVSPTERKHLTDGATVREICSAITMSDNTAANLLITITGPKETAFTHNKGDAY 134
XX |NDLYEVSPTERKHLTDGATVREICSAITMSDNTAANLLITITGPKETAFTHNKGDAY 134
XX |NDLYEVSPTERKHLTDGATVREICSAITMSDNTAANLLITITGPKETAFTHNKGDAY 134
XX
XX 135 TRLDREPELNAIPDERDRTTPVAMATT 164
XX |TRLDREPELNAIPDERDRTTPVAMATT 164
XX |TRLDREPELNAIPDERDRTTPVAMATT 164
XX
XX 61 TRLDREPELNAIPDERDRTTPVAMATT 90
XX |TRLDREPELNAIPDERDRTTPVAMATT 90
XX |TRLDREPELNAIPDERDRTTPVAMATT 90
XX

```

```

RESULT 9
ID AAB60628
XX AAB60628 standard; Protein; 159 AA.
XX
XX AAB60628;
XX
XX 25-MAR-2003 (updated)
XX 17-JUN-1991 (first entry)
XX
XX Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX MISC-difference 97..100
XX Protein 101..159
XX /label= adaptor
XX /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgaki H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
XX Matsushiro S;
XX
XX WPI; 1986-015031/03.
XX
XX N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX and transformed cells contg. 1t.
XX
XX Disclosure: Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E.coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obtd. only
XX in small ants. from human urine.
XX See also AAN60628, and 30-32.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 159 AA;
XX
XX Query Match 30.0%; Score 404; DB 7; Length 159;
XX Best Local Similarity 68.3%; Pred. No. 4,1e-34;
XX Matches 86; Conservative 9; Mismatches 19; Indels 12; Gaps 2;
XX
XX 1 HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 60
XX |HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 60
XX |HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 60
XX
XX 24 HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 83
XX |HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 83
XX |HPETLVKVDADADQCARVGYIELDNGSEILSEPRERPPMSTFVLLCGAVLSRD 83
XX
XX 61 AQOEDLGRRIHYSQNDLYEVS-----PYTEKHLDGATVREICSAITMSDNTAANL 112
XX |AQOEDLGRRIHYSQNDLYEVS-----PYTEKHLDGATVREICSAITMSDNTAANL 112
XX |AQOEDLGRRIHYSQNDLYEVS-----PYTEKHLDGATVREICSAITMSDNTAANL 112
XX
XX 84 AQOEDLGRRIHYSQNDLYEVSAAKRNDSDECPISH-----DGYCLHDSVCHYIEALDKYACNC 139
XX |AQOEDLGRRIHYSQNDLYEVSAAKRNDSDECPISH-----DGYCLHDSVCHYIEALDKYACNC 139
XX |AQOEDLGRRIHYSQNDLYEVSAAKRNDSDECPISH-----DGYCLHDSVCHYIEALDKYACNC 139
XX
XX 113 ILITIG 118
XX |ILITIG 118
XX |ILITIG 118
XX

```


XX	AAU45138	standard; Protein: 158 AA.
ID	MAU45138	
AC	AAU45138;	
DT	27-FEB-2002	(first entry)
DE	Propionibacterium acnes immunogenic protein #6034.	
KW	SAPHO syndrome; synovitis; acne; pustulosis; hypertostis; osteomyelitis;	
KM	uvellitis; endophthalmitis; bone; joint; central nervous system; ELISA;	
KW	Inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;	
KM	dematological; osteopathic; neuroprotectant.	
OS	Propionibacterium acnes.	
PX	MO200181581-A2.	
PN	01-NOV-2001.	
PP	20-APR-2001; 2001WO-US12865.	
PE	21-APR-2000; 2000US-199047P.	
PR	02-JUN-2000; 2000US-208841P.	
PR	07-JUL-2000; 2000US-216747P.	
PA	(CORI-) CORIXA CORP.	
PI	Skelly YAM, Pereing DH, Mitcham JL, Wang SS, Bhatia A;	
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;	
DR	N-PSDB; AAS59525.	
WP	WPI: 2001-616774/71.	
PT	Propionibacterium acnes polypeptides and nucleic acids useful for	
PT	vaccinating against and diagnosing infections, especially useful for	
PT	treating acne vulgaris -	
PS	Example 1; SEQ ID NO 6333; 1069pp; English.	
XX	Sequences AAU93105-AAU68017 represent Propionibacterium acnes immunogenic	
CC	polypeptides. The proteins and their associated DNA sequences are used in	
CC	the treatment, prevention and diagnosis of medical conditions caused by	
CC	P. acnes. The disorders include SAPHO syndrome (synovitis, acne,	
CC	pustulosis, hyperostosis and osteomyelitis), uvellitis and endophthalmitis.	
CC	P. acnes is also involved in infections of bone, joints and the central	
CC	nervous system, however it is particularly involved in the inflammatory	
CC	lesions associated with acne vulgaris. A method for detecting the	
CC	presence or absence of P. acnes in a patient comprises contacting a	
CC	sample with a binding agent that binds to the proteins of the invention	
CC	and determining the amount of bound protein in the sample. The	
CC	polypeptides may be used as antigens in the production of antibodies	
CC	specific for P. acnes proteins. These antibodies can be used to	
CC	downregulate expression and activity of P. acnes polypeptides and	
CC	therefore treat P. acnes infections. The antibodies may also be used as	
CC	diagnostic agents for determining P. acnes presence, for example, by	
CC	enzyme linked immunosorbent assay (ELISA).	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	atftp.wipo.int/pub/published_pcl_sequences.	
XX	Sequence 158 AA:	
Query Match	5.9%; Score 79.5; DB 22; Length 158;	
Best Local Similarity	33.3%; Pred. No. 4.9;	
Matches 26;	Mismatches 9; Indels 7; Gaps 2	
OY	110 ANLLTITGGCPETLAFILNMGDHYTRLMRWBPENLEAI PNDE RDTMTFVAATTLR--K 167 : : : : : Db 80 AMANWTSGAPRVSSASLTARCAASPVTGVPWEPSRPAT-----TTSAVLARDFRLRSGR 134	
OY	166 LITGELLTIAHQQLIDW 185	

```

Db      135 GRMGSWLTLATRIPLIRW 152
      1 : 1 : 1 : 1 : 1
RESULT 15
AAW55524
ID      AAW55524 standard; Protein; 200 AA.
XX
AC      AAW55524;
XX
DT      02-JUL-1998 (first entry)
XX
DE      H. pylori ORF 29ep10720_24433762.c3_39 cellular protein.
XX
KW      Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX      identification; binding compound; bacteria; life cycle; activator;
XX      inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS      Helicobacter pylori.
XX
PN      MO9737044-A1.
XX
PD      09-OCT-1997.
XX
PE      27-MAR-1997; 97WO-US05223.
XX
PF      06-DEC-1996; 96US-0761318.
XX      29-MAR-1996; 96US-0625811.
XX      02-APR-1996; 96US-0758731.
XX      25-OCT-1996; 96US-0736905.
XX      28-OCT-1996; 96US-0738859.
XX
PA      (ASTR ) ASTRA AB.
PI      Alm RA, Smith D;
DR      WPI: 1997-503122/46.
DR      N-PSDB: AAV24933.
XX
PT      Helicobacter pylori nucleic acid sequences and encoded
PT      polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT      infection and for diagnosis of H. pylori infection
XX
PS      Claims 14,93: Page 726-727; 1145pp; English.
XX
CC      This sequence is a H. pylori cellular protein.
CC      The protein may be used in a vaccine to prevent or treat H. pylori
CC      infection or to identify H. pylori life cycle activators or inhibitors. The
CC      useful or potential H. pylori life cycle activators or inhibitors. The
CC      DNA and probes derived from it may be used for the identification of
CC      H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC      acid sequences complementary to the DNA act as antisense sequences and
CC      can be used to prevent the translation of H. pylori mRNA. Antibodies
CC      against the protein can be used in immunoassays to evaluate the abundance
CC      and distribution of H. pylori-specific antigens. The genomic sequence of
CC      H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC      by mechanically shearing the bacterial DNA. The sequences were analysed
CC      for ORF of at least 180 nucleotides, and the predicted coding regions
CC      defined by computer evaluation. To identify likely H. pylori antigens for
CC      vaccine development, the amino acid sequences predicted from various ORF
CC      were analysed for significant homology to other known or exported
CC      membrane proteins. Having identified and determined the sequences of
CC      interest, particular regions can be isolated from H. pylori by PCR
CC      amplification for recombinant polypeptide production, e.g. in E. coli
CC      hosts.
XX
SQ      Sequence 200 AA:
XX
Query Match 5.9%; Score 79; DB 18; Length 200;
Best Local Similarity 26.0%; Pred. No. 7.8;
Matches 51; Conservative 31; Mismatches 44; Indels 70; Gaps 13
22 IELDINSGETLSEF--RPERFRPMWSTFKVLLGCAVLSRIDAGQGLGRIRIHSQNDLV 78

```



```
Db      11  VVIDTDSOKSMETFAIRAEKERPTFSLF-----NRSSGFSPTLMQWVKYENITLI 61
QY      79  ---EYSPYTEKHLTDGMTVRELCSAITMSDNTAANLL- TTIGPKEL-TAFLHNG 131
Db      62  DFKGEYSKETOK-----AMLLS-----NIVLVPTT---PSOLDTEVLANML 99
QY      132 DHYTRLDREPELNE--AIPNDERDTMFPVAMATTLKLLTGEL/TLASROOLIDME- 187
Db      100 ERIEOLQ---ELNENIRALIVINRMPTIP-----TLKERQALIEPIKE 139
QY      188 ---ADKVAGPLLRSAI 200
Db      140 NNPSDRIT--LLESSI 153
```

Search completed: September 10, 2003, 12:24:50
Job time : 33.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVG.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues
Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/ECTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	5.5	169	US-09-328-352-4491	Sequence 4491, App
2	73.5	5.5	170	US-08-858-207A-519	Sequence 519, App
3	72.5	5.4	124	US-08-311-731A-202	Sequence 202, App
4	71	5.3	159	US-08-991-890-4	Sequence 4, App11
5	70.5	5.2	197	US-09-252-991A-24241	Sequence 24241, A
6	65.5	4.9	158	US-09-010-809-19	Sequence 19, App1
7	65.5	4.9	178	US-09-252-991A-29942	Sequence 29942, A
8	64.5	4.8	198	US-09-413-814-87	Sequence 87, App1
9	64.5	4.8	203	US-09-252-991A-24921	Sequence 24921, A
10	64	4.7	150	US-09-239-909-2	Sequence 2, App11
11	64	4.7	203	US-08-624-677A-2	Sequence 2, App11
12	63	4.7	174	US-08-557-122A-6	Sequence 6, App11
13	63	4.7	174	US-09-262-666-6	Sequence 6, App11
14	63	4.7	200	US-08-557-122A-12	Sequence 12, App1
15	63	4.7	200	US-09-262-666-12	Sequence 12, App1
16	62.5	4.6	144	US-08-225-480-4	Sequence 4, App11
17	62.5	4.6	144	US-09-118-445-4	Sequence 4, App11
18	62	4.6	132	US-09-252-991A-22681	Sequence 22681, A
19	61.5	4.6	142	US-09-345-473E-31	Sequence 31, App1
20	61	4.5	199	US-09-252-991A-30363	Sequence 30363, A
21	61	4.5	204	US-09-252-991A-17611	Sequence 17611, A
22	60.5	4.5	167	US-08-690-849-2	Sequence 2, App11
23	60.5	4.5	167	US-09-004-053-2	Sequence 2, App11
24	60	4.5	158	US-09-107-532A-4218	Sequence 4218, App
25	60	4.5	184	US-09-795-926-46	Sequence 46, App1
26	59.5	4.4	103	US-09-732-210-1282	Sequence 1282, App
27	59.5	4.4	146	US-09-134-001C-5269	Sequence 5269, App

28	59.5	4.4	167	US-09-328-352-6278	Sequence 6278, App
29	59.5	4.4	171	US-09-252-991A-22351	Sequence 22351, A
30	59.5	4.4	178	US-09-220-731-24	Sequence 24, App1
31	59.5	4.4	178	US-09-242-999-24	Sequence 24, App1
32	59	4.4	136	US-09-252-991A-30622	Sequence 30622, A
33	59	4.4	144	US-09-252-991A-25578	Sequence 25578, A
34	59	4.4	148	US-09-134-001C-3640	Sequence 3640, App
35	59	4.4	171	US-09-107-532A-3979	Sequence 3979, App
36	59	4.4	180	US-09-194-905-5	Sequence 5, App11
37	59	4.4	181	US-08-482-142-195	Sequence 195, App
38	59	4.4	181	US-08-478-572-195	Sequence 195, App
39	59	4.4	181	US-08-484-296-195	Sequence 195, App
40	59	4.4	189	US-08-671-548C-48	Sequence 48, App1
41	58.5	4.3	102	US-08-479-744A-47	Sequence 47, App1
42	58.5	4.3	102	US-08-280-757B-47	Sequence 47, App1
43	58.5	4.3	174	US-09-072-596-281	Sequence 281, App
44	58	4.3	168	US-08-451-947-6	Sequence 6, App11
45	58	4.3	168	US-08-424-826A-6	Sequence 6, App11

ALIGNMENTS

```
RESULT 1
US-09-328-352-4491
; Sequence 4491, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4491
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

Query Match      5.5%; Score 74; DB 4; Length 169;
Best local similarity 23.3%; Pred. No. 3.2;
Matches 37; Conservative 34; Mismatches 62; Indels 26; Gaps 7;

OY      118 GGRKELTAFHNGDHTRLDRNEPRLNPAIDN-DEDTTPVYAMATTIRKLITG-----E 172
        ||      |      |      |      |      |      |      |      |      |
DB      5 GGEFTIMKMLAQIQN---RFQDWEQIVQYLDRLTFVRERIMVFETTFVYVIVGYSIMK 61
OY      173 LITLASROQ-----LIDWMEADKVAGPLRLSLPAGMFIADKSGAGER--GSGRITA 222
        :      |      |      |      |      |      |      |      |      |
DB      62 MHSIAEQDQRRLNDLKDLMWMMQSNV-----TKRPAELEDKSGKIORVAQDQGLTV 115
OY      223 ALGPDGPRIVYITGSOATWDERNRQIAEIGASLIK 261
        :      |      |      |      |      |      |      |      |      |
DB      116 SSGQNGEQGLTV--THQVYALANFLTQLAQLGSLIQ 152

RESULT 2
US-08-858-207A-519
; Sequence 519, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328e1 Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
```

CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 519:
SEQUENCE CHARACTERISTICS:
LENGTH: 170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-519

Query Match 5.5%; Score 73.5; DB 4; Length 170;
Best Local Similarity 22.1%; Pred. No. 3.7;
Matches 47; Conservative 32; Mismatches 71; Indels 63; Gaps 8;
QY 15 LGARGYELDLNSELLESFRPERFPMSTFVLLCGANLSIDAGQ--EQGGRIRHY 72
DB 1 LKRNIGLVLDLSIRELSPISCTKKSVMRFR-----AAAVVDASFIQELPORYD 54
QY 73 SONDLYEYPTVEKHLTDGNTVRELCSAITYMSDNTAANLLLTIGPKELTAPLHMWD 132
DB 55 -----SPYSEKSGSSTGQROLAFLARTVASQ-----PKILL----- 86
QY 133 HVRDLRHEPELANINDEMDTTPYAMATTLKLLTGELTLASNOQLDMWEADVA 192
DB 87 -----LDPATANIDSETSLV--QASLAKMKQGR-TTIAIAHRLSTIODANCI- 131
QY 193 GPLRSALPAGWFIADKSGAGERSGRTIAALG 225
DB 132 -----YLDKGRIGIESOTHELLALG 152

RESULT 3
US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: MOLE GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202

Query Match 5.4%; Score 72.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 2.9;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;
QY 130 MGDHV--TRLDREPELN---EAIPIDEDT---TPYAMATTLKLLTGELTLASRO 180
DB 12 MGDSIGMERGRNTGNTQCLRYVPGDSEPTLGRASPEDLIIT--NLSPTTMSHPPS 69
QY 181 QLDWMEA-DVAGPLI-----RSALPAGWFIADKSGAGERSGRTIAALGPKRPSR 232
DB 70 RDDDWEPFDALQTAVFATGDKATPFAVGJ-----GASTROS-GILASISPRQPAR 123

RESULT 4
US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6114307
GENERAL INFORMATION:
APPLICANT: Jaepers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:

NAME: Sawislak, Deborah A
 REGISTRATION NUMBER: 37,438
 REFERENCE/DOCKET NUMBER: 96-41
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 206-442-6672
 TELEFAX: 206-442-6678
 TELEX:
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 159 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: Internal
 US-08-991-890-4

Query Match 5.3%; Score 71; DB 3; Length 159;
 Best Local Similarity 23.8%; Pred. No. 6.3;
 Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

OY 49 VLLCGAVLSRIDAGE-----OLGRRIHYSN-----DLVEYSPTVERH 87
 DB 12 LLLCGAVF--VSPSQEIHAEFGRRHHHHGGSGAELRGCGPRGKLLSYCPMPKPT 69
 OY 88 LTDGTVRELCSAATMSDNTAANLLTTGGPKELTAFLHMGDHYTRLDRWEPELNEA 147
 DB 70 FT-----TTPGGMILES-GRPKREVSTSNKDCOAL-----GTTSEF 105
 OY 148 IPNDERDTMPVAMA-TTLRLKL 169
 DB 106 IPNLSELPKPLSEGPPLKRTI 128

RESULT 5
 US-09-252-991A-24241
 ; Sequence 24241, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 24241
 ; LENGTH: 197
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24241

Query Match 5.2%; Score 70.5; DB 4; Length 197;
 Best Local Similarity 23.6%; Pred. No. 10;
 Matches 37; Conservative 22; Mismatches 43; Indels 55; Gaps 7;

OY 94 VRELCSAATMSDNTAANLLTTGGPKELTAFLHMGDHYTRLDRWEPELNEAIPNDER 153
 DB 12 VAKLSAATMSAASLAASVPLPIATP--TSARFSAGASLT-----PSPVI 55
 OY 154 DTTMVA-MATTLRLKLTGELLTIA-----SRQQLDWMEDAVGP----- 194
 DB 56 ATTSPLACRACRTRFSLCSG--LARANTSPGSTSR--WSSNSISAPVAGSNMPT 110
 OY 195 -----LMSALPAGWFIADKSGAGRG 216
 DB 111 SMAPLAAVSTWSPVITFTAIAPAAHSAATTAATASRG 147

RESULT 6
 US-09-010-809-19
 ; Sequence 19, Application US/09010809B
 ; Patent No. 6090601
 ; GENERAL INFORMATION:
 ; APPLICANT: Gustafson, Claes
 ; APPLICANT: Bellach, Mary C.
 ; TITLE OF INVENTION: Epothilone Polyketide Synthases and Encoding DNA
 ; FILE REFERENCE: 30062-20020.00
 ; CURRENT APPLICATION NUMBER: US/09/010,809B
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 19
 ; LENGTH: 158
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-010-809-19

Query Match 4.9%; Score 65.5; DB 3; Length 158;
 Best Local Similarity 27.6%; Pred. No. 25;
 Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

OY 106 DNTAANLLT-----TIGPKELTAFLHMGDHYT--RLDRWEPE-----ELNEAIPNDERD 154
 DB 21 NHDAKHLITSRQASAPGADVRLSELALGASVTLAAGDVADPRALRDLNDIP----- 75
 OY 155 TTMVAMATTLRLKLITGELLTLASRQQLIDMNEADKRVAGPLRSA-----LPAGWF 205
 DB 76 SAHPAAVVAHVAASTVDGDLGAMSLERI-----DRVFAPKIDAAWHQLTQDRPLAAF 129
 OY 206 IADKSGAGRGSGRG 219
 DB 130 ILFSSVAGVIGSSG 143

RESULT 7
 US-09-252-991A-29942
 ; Sequence 29942, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 29942
 ; LENGTH: 178
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-29942

Query Match 4.9%; Score 65.5; DB 4; Length 178;
 Best Local Similarity 26.8%; Pred. No. 30;
 Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

OY 110 ANLLTTG--GPELTAFA-----LHMGDHYTRLDRWEPELNEAIPNDERDTMPVAMA 162
 DB 25 AVLLVEALQORRPELAAHAGLAHHPCDHYAR-----AGEDLP--RDVQLGAAA 74
 OY 163 TILRLKLTGELL-----TLASRQQLIDMNEADKRVAGPLRSAIPAGWFIADKSG 211

```

Db 75 LRHHKLYAGALGSEHHNEEGISDGTAAEGDAVVG--QDREVAGRPV--GLDGLGELTAMVGD 130
QY 212 AGERGSRGIIAALGPRDK 229
Db 131 A-----LVGVVGQAQR 141

```

```

RESULT 8
US-09-413-814-87
: Sequence 87, Application US/09413814
: Patent No. 6225064
: GENERAL INFORMATION:
: APPLICANT: Geesellschaft fuer Biotechnologische Forschung mbH
: APPLICANT: Bristol-Myers Squibb, Co.
: APPLICANT: Beyer, Stefan
: APPLICANT: Bloecker, Helmut
: APPLICANT: Brandt, Petra
: APPLICANT: Cino, Paul M
: APPLICANT: Dougherty, Brian A
: APPLICANT: Goldberg, Steven L
: APPLICANT: Hofte, Gerhard
: APPLICANT: Mueller, Joachim
: TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
: FILE OF INVENTION: heteropolypeptide compounds
: FILE REFERENCE: PCT/US 99/23535
: CURRENT APPLICATION NUMBER: US/09/413,814
: CURRENT FILING DATE: 1999-10-07
: EARLIER APPLICATION NUMBER: DE 198 46 493.2
: EARLIER FILING DATE: 1998-10-09
: NUMBER OF SEQ ID NOS: 107
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 87
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
: US-09-413-814-87

```

Query Match	4.8%	Score 64.5	DB 3	Length 198
Best Local Similarity	22.4%	Pred. No. 46		
Matches	62	Conservative	30	Mismatches 78; Indels 107; Gaps 15
QY	1	HPETLVKKQAEQDQGARVGYIELDNSGEILSEFRPERPMSTF----	KYLLGAV	55
DB	7	NPEAVNDVAKDKTS---AAFGV---DATVHKKLEGICAQYETAIVTAEHNEPSVLCVAS	60	
QY	56	LSRIADAGEQLGRRIHYQSNDLVEFSPTTEKHLTDGAVRELGSAITMSDNTAAILLT	115	
DB	61	LVR-----KGRITIAAA---TBROAD--TLR	82	
QY	116	TIGSPKELTAFLNHMGCHVTRLDNRPEPLNEAIPDERDTHPYAAMATTLKLTGELLT	175	
DB	83	AVG-----ATRVQLDETEMGRRV---GADITMPLAQ-----DLDD	114	
QY	176	LASRQQLDMMEDAKVAGPILRSALPGWFTADK---SGAGER-----GSRGIITALG	225	
DB	115	LASHYRYVPM-----NAHGPLVAGQTL-AQSKTRQGYRIYVLGVRPHTRNRPDDKFRLEAPY	169	
QY	226	PDGKPSRIIVAYITTTGSAQTM-----DERNQIAEIG	256	
DB	170	PD-----YVIRGDGTLTLLVAGSDSDVSREVAEVG	197	

```

RESULT 9
US-09-252-991A-24921
: Sequence 24921, Application US/09252991A
: Patent No. 6531795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERGININOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252.991A

```

```

1  CURRENT FILING DATE: 1999-02-18
2  PRIOR APPLICATION NUMBER: US 60/074,788
3  PRIOR FILING DATE: 1998-02-18
4  PRIOR APPLICATION NUMBER: US 60/094,190
5  PRIOR FILING DATE: 1998-07-27
6  NUMBER OF SEO ID NOS: 33142
7  SEO ID NO 24921
8  LENGTH: 203
9  TYPE: PRT
10 ORGANISM: Pseudomonas aeruginosa
11 US-09-255-991A-24921

```

```

Query Match      4.8%; Score 64.5; DB 4; Length 203;
Best Local Similarity 27.1%; Pred. No. 46;
Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7.

QY 135 TRLDREPEL-----NEAIPNDERDITMEVAAATTLRKILITGELLTASROOLIDME 187
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 34 TPLDSSERLNVSLAWRKAMPRPVQAORMPKLAIEI--RALTGLSTLTSSR---DWS- 86
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 188 ADKVAGPLLRNALPAGWFIADKSCAGSGESKIIIALCPDGKPSRIIVITYITGSGATDE 247
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 87 -----TSAMPGGCDI-----SPGWLASTLEGR-AKLAITATFT---AALKV 123
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 248 RNRLAELG 256
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 124 PTVYSTTG 132
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

RESULT 10
US-09-239-909-2
: Sequence 2, Application US/09239909
: Patent No. 6284952
: GENERAL INFORMATION:
: APPLICANT: Kumbo Petrochemical Co. Ltd.
: TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
: FILE REFERENCE: P99-2-6
: CURRENT APPLICATION NUMBER: US/09/239,909
: CURRENT FILING DATE: 1999-01-29
: EARLIER APPLICATION NUMBER: EP 99300136.1
: EARLIER FILING DATE: 1999-01-08
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: KOPATIN 1.0
: SEQ ID NO 2
: LENGTH: 150
: TYPE: PRT
: ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

```

```

Query Match Similarity      4.7%; Score 64; DB 3; Length 150;
Best Local Similarity      27.9%; Pred. NO. 34;
Matches      36; Conservative 17; Mismatches 60; Indels 16; Gaps 6

QY      73 SQNDLVESPYTEKILTDG---MIVRELCSAITYSDNTAANLLTTI-----GPKE 122
       ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      6 SEQGIYDPRKEAFGLELDKDGDCGICITVELATVIRSLDQNPTREELDDMSIENVADNGTIE 65
       ||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      123 LTAFLHNKGHWTRIDRWEPELEINAIIPNDERDTTPVAMATTLRKL--TGELLTLASRQ 180
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      66 FDEELSTLAKRKVKDTDA-EEELKEAFKFKDKDONGYIS-ASELRHVMIMLSKELTDEVEYE 123
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY      181 QLIOMMEAD 189
       ||::|||
Db      124 QMI--KEAD 130

```

RESULT 11
US-08-624-677A-2
; Sequence 2, Application US/08624677A
; Patent No. 646192
; GENERAL INFORMATION:
; APPLICANT: Tally, Nicola C.
; APPLICANT: Jenkins, Mark C.

APPLICANT: Dubey, Jitender P.
TITLE OF INVENTION: Antigens Useful for the Serodiagnosis of
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Graeter, Janelle S.
STREET: Rm. 411, Bldg. 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,677A
FILING DATE: 15-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0228.95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-6629
TELEFAX: 301-504-5060
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-677A-2

Query Match 4.7%; Score 64; DB 4; Length 203;
Best Local Similarity 23.8%; Pred. No. 55;
Matches 39; Conservative 23; Mismatches 72; Indels 30; Gaps 7;

QY 54 AVLSRIDAGQEQQLGRRIHYSNDLVEYSPVTEKHLDGTVRELCSAA--ITMSDNTAAN 111
DB 33 AGSVNYDDGDMDAAGNPVD-----SDVTDAITDGEWPRVVSQKPHHTQKSLIAK 83
QY 112 LLLTTIGPKETATLHNKGHY-TRLDWPEELNKAIPNDR-DTTPVA----- 160
DB 84 IAVPVGA--LTSYL-VADRYLDELTSAEEGTESIPGKRKYTAVGIALVAANA 138
QY 161 ---MATTLRLKLTGELLFLASROQLDWEADKVAAGPLRSALP 201
DB 139 GIGLARTFRHVPKRSKIVASDASLGNSEQYEGTVAGSSDP 182

RESULT 12
US-08-557-122A-6
Sequence 6, Application US/08557122A
Patent No. 5879664
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Molland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5879664 of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-6

Query Match 4.7%; Score 63; DB 2; Length 174;
Best Local Similarity 23.9%; Pred. No. 56;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KYKDAEDQGAR-VGIIEDLNSGETLESFRRERPMKSTKYLICGAVLSRIDAGQEQ 65
DB 67 KYEQATELKERINIPLVKDYCTEEBALCRDQVEGYPTLKIRGL-----DAVKRY 117
QY 66 LGRR-----IHYS-QNDLVEYSPVTEKHLDGTVRELCSAAITMSDNTAANLLLT 116
DB 118 QGARQTEALVSTMYKQSLPRAVSPVTPENLEIKTKMDKIYVIGIASDDQTANDIFFT 174

RESULT 13
US-09-262-666-6
Sequence 6, Application US/09262666
Patent No. 6346244
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Molland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: No. 6346244 of No. 6346244th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,666
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,122
FILING DATE: 11-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-262-666-6

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 63462440 No. 63462444disk of No. 63462444th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,666
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,122
FILING DATE: 11-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980,204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-262-666-12

Query Match          4.7%; Score 63; DB 4; Length 200;
Best Local Similarity 23.9%; Pred. No. 69;
Matches 26; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KVDAEDQDGLAR-VGYIELDLNNGEILSEFRPERFPMMSTFKVULGSAVLRIIDAGQEQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 43 KYEQAATELKEKNKPIPLKVKVDTSEALCRDQGVGGYPLKIFRGL-----DAVKPY 93

QY 66 LGRR-----IHS-QNDLVEISPYTEKHLDGATVRELCSAATMSDNTAANLLFTT 116
   | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 94 QGARQTEALIVSYWVKOSLPAYSPYTPENLEIKTMKRIKIVIGYIASDDQTDNDIFTT 150

Search completed: September 10, 2003, 12:31:02
Job time : 12.8571 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(Without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E
Sequence: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVG.....TNDERKRLQIAELIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgnt2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgnt2_6/ptodata/2/pubpaa/PCMT_NEM_PUB.pep.*
3: /cgnt2_6/ptodata/2/pubpaa/US06_NEM_PUB.pep.*
4: /cgnt2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgnt2_6/ptodata/2/pubpaa/US07_NEM_PUB.pep.*
6: /cgnt2_6/ptodata/2/pubpaa/PCMTUS_PUBCOMB.pep.*
7: /cgnt2_6/ptodata/2/pubpaa/US08_NEM_PUB.pep.*
8: /cgnt2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgnt2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgnt2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgnt2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgnt2_6/ptodata/2/pubpaa/US09_NEM_PUB.pep.*
13: /cgnt2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgnt2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgnt2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgnt2_6/ptodata/2/pubpaa/US10_NEM_PUB.pep.*
17: /cgnt2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
18: /cgnt2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	463	34.3	94	US-10-102-806-760	Sequence 760, App
2	78	5.8	202	US-10-127-816-9	Sequence 9, Appl
3	78	5.8	202	US-10-142-717-12	Sequence 12, Appl
4	69.5	5.2	149	US-10-233-926-4	Sequence 4, Appl
5	69	5.1	149	US-10-156-761-8136	Sequence 8136, Ap
6	69	5.1	178	US-09-738-626-5880	Sequence 5880, Ap
7	69	5.1	206	US-10-156-761-13867	Sequence 13867, A
8	67.5	5.0	184	US-10-101-464A-765	Sequence 765, App
9	67.5	5.0	187	US-09-948-018-8	Sequence 8623, Ap
10	67	5.0	194	US-10-156-761-8623	Sequence 8623, Ap
11	65.5	4.9	191	US-10-127-816-11	Sequence 11, Appl
12	65.5	4.9	191	US-09-764-870-390	Sequence 390, App
13	65.5	4.9	191	US-10-125-540-390	Sequence 390, App
14	65	4.8	179	US-09-764-868-757	Sequence 757, App
15	65	4.8	179	US-10-106-698-4858	Sequence 4858, Ap

16	65	4.8	202	US-10-189-346-12	Sequence 12, Appl
17	64.5	4.8	193	US-10-156-761-8978	Sequence 8978, Ap
18	64.5	4.8	193	US-09-805-354-8	Sequence 8, Appl
19	64.5	4.8	193	US-10-144-259-8	Sequence 8, Appl
20	63.5	4.7	189	US-09-815-242-5862	Sequence 5862, Ap
21	63.5	4.7	189	US-09-815-242-12979	Sequence 12979, A
22	63.5	4.7	206	US-09-738-626-5425	Sequence 5425, Ap
23	62.5	4.6	144	US-10-131-406-4	Sequence 4, Appl
24	62.5	4.6	190	US-10-156-761-19507	Sequence 9507, Ap
25	62.5	4.6	195	US-10-156-761-12656	Sequence 12656, A
26	62	4.6	152	US-09-738-626-4796	Sequence 4796, Ap
27	62	4.6	174	US-09-864-761-35777	Sequence 35777, A
28	62	4.6	190	US-09-738-626-4637	Sequence 4637, Ap
29	62	4.6	202	US-10-156-761-13460	Sequence 13460, A
30	61.5	4.6	88	US-10-100-252-6	Sequence 6, Appl
31	61.5	4.6	142	US-09-862-027-31	Sequence 31, Appl
32	61.5	4.6	187	US-10-156-761-12111	Sequence 12111, A
33	61.5	4.6	196	US-09-738-626-6624	Sequence 6624, Ap
34	61.5	4.6	202	US-10-189-346-16	Sequence 16, Appl
35	61	4.5	70	US-09-864-761-41555	Sequence 41555, A
36	61	4.5	177	US-09-791-9937-69	Sequence 69, Appl
37	61	4.5	203	US-09-800-729-154	Sequence 154, App
38	60.5	4.5	152	US-10-156-761-14346	Sequence 14346, A
39	60.5	4.5	160	US-09-882-227-230	Sequence 230, App
40	60.5	4.5	196	US-10-219-220-265	Sequence 265, App
41	60.5	4.5	200	US-10-156-761-12086	Sequence 12086, A
42	60	4.5	134	US-09-768-2358-40	Sequence 40, Appl
43	60	4.5	184	US-09-798-029-4	Sequence 4, Appl
44	60	4.5	184	US-09-795-926-46	Sequence 46, Appl
45	60	4.5	184	US-10-364-774-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103PIC1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-06-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-760

Query Match
Best Local Similarity 34.3%; Score 463; DB 15; Length 94;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
75 NDLYEYSPVTEKHLIDGKTVRELCSAATMSDNTANLTLTTGGPRELTAFLNMGHV 134
DB 1 NDLYEYSPVTEKHLIDGKTVRELCSAATMSDNTANLTLTTGGPRELTAFLNMGDGV 60


```
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIRAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHITAKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8136
LENGTH: 149
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8136
```

```
Query Match 5.1%; Score 69; DB 15; Length 149;
Best Local Similarity 30.9%; Pred. No. 34;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;
```

```
OY 168 LTTGELT-----ASROQLDMMEDAVAGPLRLSALPAGWFIADKSG----- 211
Db 50 LDIETLTJTKRLRLVASVDKAKEMGIDWMEHD-----PALSSHADGGRRLAEKRLREI 105
OY 212 AGERSGRTIATGPGKPSR 232
Db 106 AGLRQAALPSAGPAKPSR 126
```

```
RESULT 6
US-09-738-626-5680
Sequence 5680, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5680
LENGTH: 178
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680
```

```
Query Match 5.1%; Score 69; DB 10; Length 178;
Best Local Similarity 22.1%; Pred. No. 44;
Matches 34; Conservative 19; Mismatches 65; Indels 36; Gaps 5;
```

```
OY 113 LTTGCPKELTAFELNMGDH--VTRLDMEPELNEAIPNDEDTTTPVAMATTLRLLT 170
Db 27 LKVTKAGPKSAVALIKVDSRPLDQLEVSQSIGELFQDAEDRGELNFGAGTTLVSTP 86
OY 171 GELLTASRQQLDMMEDAVAGPLRLSALPAGWFIADKSGAGERSGRTIATGPGK 230
Db 87 G-----VDNPL--TLPRM-----RRNRGRVALDDGKK 114
OY 231 --SRIYVITTSQATMDERNQIATIGASLIRH 262
Db 115 RVARIGALNDAETHVLIERNKLEVTTLLELAH 148
```

```
RESULT 7
US-10-156-761-13867
Sequence 13867, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIRAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHITAKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13867
LENGTH: 206
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13867
```

```
Query Match 5.1%; Score 69; DB 15; Length 206;
Best Local Similarity 25.2%; Pred. No. 55;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;
```

```
OY 70 IHYSQNDLVEYSVTEKHLTDGMTV-----RELCSNAITMSDNTAAILTTI---G 118
Db 1 VHASQG---AQGGSAGEKEKVTMDAAGSESEFEVA-----NRSSALLTAVLLSG 50
OY 119 GPKR---LTAFLHNNGDHVRILD-----RM-----EPDLNEAIPNDE 152
Db 51 GDRHAAEDLDQNALIKRADRSRIDPEAVVQVLRQVSRWRKWRRELSVAEP-PE 109
OY 153 RDTTTPVAMATTLRLKLTGELLTASRQQL-----DMMEDKVAAGPLRLSALPAGW 205
Db 110 ASTGPDAASAEELVWAGALARIATANQVTLVRYFEDLPDAD--VARITGSGVTVRS 167
OY 206 IADSGAGERSGRTIATGPG---DGKPSR 232
Db 168 TTHRSLARLRTLAPDLAALGPAADGDEPSR 197
```

```
RESULT 8
US-10-101-464A-765
Sequence 765, Application US/10101464A
Publication No. US20030046728A1
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions isolated from Plant Cells
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
FILE REFERENCE: 11000.1020C2
CURRENT APPLICATION NUMBER: US/10/101,464A
```



```
RESULT 15
US-10-106-698-4858
; Sequence 4858, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OR INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4858
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo.sapiens
US-10-106-698-4858

Query Match          4.88; Score 65; DB 15; Length 179;
Best Local Similarity 27.3%; Pred. NO. 1.2e+02;
Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

QY 120 PKELTATLHMGDHVRD---RWEPELNEA-----IPNDERDTMPVAMATIRKLLTG 171
   |||||  ||: || || |||||  ||:: |  |
DB 52 PRELTIVY---QGEKLEVLDHRSKRMMLVKNEAGRSGLTIPSNILEPLQPGTPTGQGPSRV 108
   |||||  ||: || || |||||  ||:: |  |
QY 172 ELLTLASR-QQLIDMEADKVAGPLRS 198
   |||||  ||: || || |||||  ||:: |  |
DB 109 PMLRLSSRPEVTDWLQANFNSTATVRT 136
   |||||  ||: || || |||||  ||:: |  |
```

Search completed: September 10, 2003, 12:33:17
Job time : 17.5714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPEITLVKVKDAEDQAGARVC.....TMDERNRQIAETGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues
Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	534	38.9	105 2	JC2566 bla protein - pseu
2	78	5.8	191 2	hypothetical prote
3	75.5	5.6	190 2	ADP-ribosylation f
4	75.5	5.6	200 2	spore coat protein
5	74	5.5	177 2	hypothetical prote
6	73.5	5.5	195 2	hypothetical prote
7	72.5	5.4	113 2	hypothetical prote
8	71.5	5.3	152 2	hypothetical prote
9	71	5.3	192 2	conserved hypotet
10	71	5.3	198 2	conserved hypotet
11	70.5	5.2	131 2	hypothetical prote
12	70.5	5.2	184 2	ADP-ribosylation f
13	70	5.2	108 2	hypothetical prote
14	69.5	5.2	145 2	hypothetical prote
15	68.5	5.1	172 2	conserved hypotet
16	68.5	5.1	172 2	molycoprotein bios
17	68.5	5.1	180 2	hypothetical prote
18	68	5.0	170 2	conserved hypotet
19	67.5	5.0	192 2	hypothetical prote
20	67.5	5.0	198 2	conserved hypotet
21	66.5	4.9	128 2	conserved hypotet
22	66	4.9	149 2	nucleoside-diphosp
23	65.5	4.9	42 2	collagen alpha 1(X
24	65.5	4.9	116 2	ribosome binding f
25	65.5	4.9	195 2	probable bacterioph
26	65	4.8	146 2	hypothetical prote
27	65	4.8	148 2	conserved hypotet
28	65	4.8	150 2	probable heat choc
29	65	4.8	153 2	conserved hypotet

30	65	4.8	160 1	E69186 conserved hypotet
31	65	4.8	168 2	B75498 conserved hypotet
32	65	4.8	180 2	G70912 hypothetical prote
33	64.5	4.8	151 2	probable protein-t
34	64.5	4.8	177 2	D81333 hypothetical prote
35	64	4.7	113 2	D70580 hypothetical prote
36	64	4.7	150 2	T08585 hypothetical prote
37	64	4.7	162 2	AG0769 calmodulin - soybe
38	63.5	4.7	168 2	T20606 probable acetyltra
39	63.5	4.7	177 2	S11602 hypothetical prote
40	63.5	4.7	179 2	AB1994 hypothetical prote
41	63.5	4.7	182 2	A96202 nwsb protein (limp
42	63.5	4.7	196 2	H69647 2-dehydro-3-deoxy-
43	63	4.7	122 2	AC3609 transcription regu
44	63	4.7	124 2	F75373 conserved hypotet
45	63	4.7	168 2	AG2257 hypothetical prote

ALIGNMENTS

RESULT 1
JC2566
bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa
C:Date: 26-Jun-1995 #sequence_rev15ion 14-Jul-1995 #text_change 03-May-1996
C:Accession: JC2566
R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148(128), 81-86, 1994
A:Title: Construction of Improved Escherichia-pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A:Note: due to a typographical error the volume number 148 appears as 128
A:Accession: JC2566
A:Molecule type: DNA
A:Residues: 1-105 <MES>

C:Genetics:
A:Gene: bla
C:Superfamily: beta-lactamase I

Query Match 38.9%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 3.4e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMTTLKLLTGELTLASRQQLIDMMEADKVGALLRSALPAGWFIADKSGAGERSRG 219
Db 2 AAMTTLKLLTGELTLASRQQLIDMMEADKVGALLRSALPAGWFIADKSGAGERSRG 61
QY 220 ITAALGPDGKPSRIIVYITGSGATMDERNRQIAETGASLIKHW 263
Db 62 ITAALGPDGKPSRIIVYITGSGATMDERNRQIAETGASLIKHW 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 20-Jul-1996 #sequence_rev15ion 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: T38062; S67447
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996

A:Reference number: Z21766
A:Accession: T38062
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <MC2>
A:Cross-references: EMBL:269944; NID:q1217974; PIDN:CAAS3808.1; PID:q1217978; GSPDB:G
A:Experimental source: strain 97zh-; cosmid c1F12
C:Genetics:
A:Gene: SPAC1F12.04c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.8%; Score 78; DB 2; Length 191;

RESULT 6
T36975
hypothetical protein SCJ11.04 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36975
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36975
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-195 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52889.1; GSPDB:GNO0070; SCODEB:SCJ11.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCJ11.04

Query Match 5.5%; Score 73.5; DB 2; Length 195;
Best Local Similarity 17.0%; Pred. No. 71;
Matches 40; Conservative 43; Mismatches 77; Indels 75; Gaps 9;
QY 15 LGARVGYIEL-DLNSEILSEFPEREPFPMSTFKVLLCGAVLSRIDAGOEGRHHS 73
D 17 MASMLGLLAREASAREVLEF-----AARAVALEAGETELDRV--- 60
QY 74 QNDLVESPVTEKHLDGNTVRELCSAATMSDNTANLLTTIGGPKELTAFLHNGDH 133
D 61 -----IARELVEALVSAEFTGTAEAGECEPALVPAPASAEFGAL----- 104
QY 134 VRLDWEPELNAI--PDERDTMPVAMATTLRLKLLIGELLTLASROOLIMMEADKY 191
D 105 ---VPMQECLSVLSVSNQR-----ILNVLDRLRGLEPVRAKDI 142
QY 192 AGFLRSALPA-----GMFIADKSGAGERSGRTIATGPDGRPS 231
D 143 AALGIEAAMAARKEVGRPKRLAERGMVLQEAAGFSAGR--LVAS--PGDPS 194

RESULT 7
T45195
hypothetical protein u1756 [Imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z16911
A:Accession: T45195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

Query Match 5.4%; Score 72.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 41;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;
QY 130 MGDHV--TRLRMEPELN--EAIPIIDERDT---TMPVAMATTLRLKLLIGELLTLASRQ 180
D 1 MGGGIGMERBGRNTGTGCPRLRVPGDESTTLGGRASPELIT--NLSPITKSHPPPS 58
QY 181 QLTIDMEA-DKVAAGPLL-----RSALPAGWFIADKSGAGERSGRTIATGPDGRPSR 232
D 59 RDDDWEPFDALOGTAVFATGDKATMPAVAGI---GASTRGS-GILASTLSPFQPAR 112

RESULT 8
T36984
hypothetical protein SCJ11.13 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000

C:Accession: T36984
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36984
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-152 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52898.1; GSPDB:GNO0070; SCODEB:SCJ11.13
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCJ11.13
C:Superfamily: Streptomyces coelicolor hypothetical protein SC9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
Best Local Similarity 23.4%; Pred. No. 74;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;
QY 106 DNTANLLTTIGGPKELTAFLHNGDHTVTRLDRMEPELNAIPNDERDTMPVAMATTL 165
D 25 DQAAARTLAQVTPPADLAFYESIGD-VT---WEDVNGYPLNP----- 65
QY 166 RLLTGEELTLASROOLIMMEADKYAGPILRSALPAGWFIADKSGAGERSGRTIATG 225
D 66 ---AGDLRLRLQEGYGVDTGDEKSRGLVI-----GSRNG-GLIYVAG 104
QY 226 PDGKPSRIYVYTTGSOATMDERN-----ROI AEI 255
D 105 PDG-----AVYRTR-ASLDEALDKVADGURQPLEL 135

RESULT 9
G83096
conserved hypothetical protein PA4399 [Imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
C:Accession: G83096
R:Stover, C.K.; Pham, X.Q.; Ewlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.D.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folgar, K.R.; Kas, A.; Lartig, K.; L.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A:Reference number: A82850; MIMD:20437337; PMID:10984043
A:Accession: G83096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <STO>
A:Cross-references: GB:AE004855; GB:AE004091; NID:G9950621; PIDN:AA607787.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA4399
C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
Best Local Similarity 26.0%; Pred. No. 11e+02;
Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;
QY 53 GAVLSRI-----DAGEQL--GRIHYQNDLVEYSPTEKHLDGNTVRELCSA----- 100
D 2 GNLSTIYRTGREGTGLAGRRVRSRIRIATGAVDELNSQLLAELEAAGAMP 61
QY 101 -----ATMSDNTANLLTTIGGPKELTAFLHNGDHTVRL---DRMEPEL-----N 145
D 62 GLREIVQALAPVHR-----LFDLGGELAMPYRAIDETEVANLESCIDRWDELPLKN 116
QY 146 EAIPINDERDTMPVAMATTLRLKLLIGELLTLASROOLIMMEADKYAGPILRSALPAGW 205
D 117 FILPGSR---PVAQAHVCRSLAR---SAERRCALDOEETLEGVGRLYINRLSDLE 168
QY 206 IADKSGAGERSGRTI--AALGPD 227
D 169 VAAARATARRGVAEIILMEAAAKPD 192

C;Genetics:
A;Gene: VNG0594H

Query Match	5.2%	Score 70:	DB 2:	Length 108;
Best Local Similarity	27.5%	Pred. NO. 62;		
Matches 28:	Conservative 16;	Mismatches 42;	Indels 16;	Gaps 3

```

QY 44 M5PEKVLICGAVLSRIDAGQSLGRIRIYSONDLVEYSPMEKHTDQMTVRELCSAII 103
Db 1 MDTVTVELDGAFLARLEATD-----RVEEVAFFDALEVTVDLRFHHDDRV-----GSII 50
QY 104 MSDMTANLILTTIGGPREL-----TAFLLNMGDHYRLDR 139
Db 51 NDDGTDRTMARLTVGGDSDFTAVEPYPSLVAIVDAARTRR 92

```

RESULT 14
F84251

hypothetical protein Vng0953c [Imported] - Halobacterium sp. NRC-1
C:Snecles: Halobacterium sp. NRC-1

hypothetical protein Vng0953c [Imported] - Halobacterium sp. NRC-1
C:Snecles: Halobacterium sp. NRC-1

C:\Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:\Accession: FA4351

RECEIVED: FEBRUARY 1964

Jung, K.H.; Alam, M.; Freltas, T.
Proc Natl Acad Sci U S A 87 12175-12181 1990

110C: Natl. Acad. Sci. U.S.A. 91, 12170-12181, 2000

A Title: Genome sequence of *Halobacterium* species NRC-1.

A; Reference number: AB4160; MUID:20504463; PMID:11016950

A: Accession: F84251

A;Status: preliminary

A;Molecule type: DNA

A;ResIdues: 1-145 <ST

A:Cross-References: GB:AE004437. NID:nl0580510. PIDN:AACT19379 1. CENDP:CV00138

C. Genet Inc.

С.В. КОЗЛОВ. УЧЕБНИК

Query Match	5.2%	Score 69.5;	DB 2;	Length 145;
Best Local Similarity	20.3%	Pred. NO.1e+02;		
Matches	30;	Conservative	29;	Mismatches 66;
				Indels 23;
				Gaps 3;

```

OY      41  EPMASTRVLLCGAVLS-----RLDAGQELGGRIRHSQNDLYEVSYTEKH  87
Db      3  FVSAGSAALFLFAFVSGLIYSAAFNGFEFVODANNARNDRVLAKTKVVEAVATTYDS  62
OY      88  LTDMFTRFELCSAITMSDNTAAILLTITIGPKELFAFLHMGDHYRILDRWPELWEA  147
Db      63  VNDIVTAVNAINNGSTLLSVSQTDVLY--DGEYVDSAYVSSVDGNSQTDILM-----  112
OY      148  IPNDEKDTMPVAMATTLKLLTGEELT  175
Db      113  LPGETYSVTVPTGSAPTFRVKVVTGTGTV  140

```

RESULT 15
D87360

conserved hypothetical protein CC0895 [Imported] - *Caulobacter crescentus*

C;species: Caulodacter crescentus

C;Date: 20-Apr-2001 #sequence_revison 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: D87360

R: Nieleman, W.C., Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Elsen, J.; Heldberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolton
J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

PROC. NATL. ACADE. SCI. U.S.A. 98, 4136-4141, 2001

A; Title: Complete Genome Sequence of *Caulobacter crescentus*

A; Reference number:

A:Accession: D87360

A: Status: preliminary

Abstract: *Phytophthora blight* caused by *Phytophthora blanda* was reported for the first time in China. The pathogen was isolated from diseased plant tissues and identified by morphological and molecular biological methods. The pathogen was found to be highly virulent to *Phytolacca esculenta* and *Phytolacca americana*. The pathogen was also found to be highly virulent to *Phytolacca* spp. in the field. The pathogen was found to be highly virulent to *Phytolacca* spp. in the field.

h, molecule type: DNA

A; Residues:

A;CROSS-TELETYPE

Query Match	5.1%	Score 68.5	DB 2	Length 167
Best Local Similarity	25.4%	Pred. No. 1.5e+02		
Matches 36, Conservative	19	Mismatches 56	Indels 31	Gaps 6

```

OY      2 PELYKYVDA-----EDQGARQYIELDN---SGELESFPEREPFMSTFK 48
        |||::|||      :|||::      :|||::      :|||::
DB      24 PWTLLILNDAGVAKRFEOQMODRLGVAARNVLAARLKLTVLSHQWMTREYSR-PPRHETW 82
        |||::|||      :|||::      :|||::      :|||::
OY      49 VLLGAVLSRIDAGEQGLGRRIHYSQNDLVESPYTEKHLLDGMIVR-----ELCSAAT 103
        |||::|||      :|||::      :|||::      :|||::
DB      83 LTEKGLASPVLTLTMAEMGDRHHYGR---DKSPVLFRRHKTCGCAFHRYPLACEAGQYVD 138
        |||::|||      :|||::      :|||::      :|||::
OY      104 MSDNTAAVILLTTTGGPRELTA 125
        |||::|||
DB      139 RRDIERA-----GPDUTA 152

```

Search completed: September 10, 2003, 12:26:15
Job time : 11.2857 secs

THIS PAGE BLANK (USPTO)

GenCode version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E

Sequence: 1 HPEITLVKVADEQDQAGRVG.....TMDERNRQIAETGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	5.8	191	YDA4_SCHPO	Q10346 schizosacch
2	71	5.3	198	OGG1_ARCFU	029876 archaeoglob
3	69.5	5.2	196	RAC2_LOTJA	040220 lotus japon
4	69	5.1	201	RACG_DICDI	098950 dictyostell
5	67	5.0	122	HC3L_THIFE	P80509 thlobacilli
6	66	4.9	149	NDK_TREPA	083374 treponema p
7	65.5	4.9	116	RBR4_UREPA	098940 ureaplasma
8	65.5	4.9	206	KTHY_METAC	081hs9 methanosarc
9	63.5	4.7	196	ALRH_BACSU	P50846 b kgp/kdp
10	63.5	4.7	200	RR4_PELNE	0946d9 pellicia nees
11	63	4.7	176	HSIV_THEMA	P27245 thermotoga
12	62.5	4.6	144	MARR_ECOLI	P27245 escherichia
13	62.5	4.6	172	YDEJ_ECOLI	P31131 escherichia
14	62.5	4.6	182	PIRE_STRCO	09487 streptomyces
15	62.5	4.6	195	TRPF_THEVO	0979v6 thermoplasma
16	62	4.6	174	IHBH_RAT	P17491 rattus norv
17	62	4.6	182	YB61_BIFLO	086520 bifidobacte
18	61.5	4.6	184	YB61_SCHPO	0949g5 schizosacch
19	61.5	4.6	185	YCCG_ECOLI	093615 xenopus lae
20	61.5	4.6	197	HANI_XENLA	073615 xenopus lae
21	61	4.5	177	VNSC_RINDR	P35948 rinderpest
22	61	4.5	177	VNSC_RINDR	003339 rinderpest
23	61	4.5	202	COAT_ELV	P35927 erythrinum la
24	60.5	4.5	126	YGM1_YEAST	P53130 saccharomyc
25	60.5	4.5	142	YB63_METTM	050770 methanobact
26	60.5	4.5	146	H8G_RABIT	P02099 oxytolaqus
27	60.5	4.5	150	PDUV_SALTY	025700 helicobacte
28	60.5	4.5	160	TATB_HELPY	025700 helicobacte
29	60.5	4.5	175	Y581_AQUAE	068848 aquifex aeo
30	60.5	4.5	178	HSIV_RALSO	0833d7 ralsonta s
31	60	4.5	121	SECR_HUMAN	P09383 homo sapien
32	60	4.5	159	GREX_CHLLE	083853 chlorobium
33	60	4.5	178	PYRE_ARCFU	028533 archaeoglob

34	60	4.5	178	1	UCRI_ANASP	P70758 anabaena sp
35	60	4.5	184	1	ARI2_DRONE	006849 drosophila
36	60	4.5	200	1	TATB_CADUC	09a6t1 caulobacter
37	59.5	4.4	103	1	RS10_NEIGO	P48851 neisseria g
38	59.5	4.4	152	1	YU33_YERPE	082cf8 yersinia pe
39	59.5	4.4	156	1	BFR_AZOVI	P22759 azotobacter
40	59.5	4.4	177	1	ATPD_HAEIN	P43717 haemophilus
41	59.5	4.4	184	1	HRPL_PSESY	P37929 pseudomonas
42	59.5	4.4	198	1	UPP_ARATH	065583 arabidopsis
43	59.5	4.4	205	1	RS4_ECOLI	P02354 escherichia
44	59.5	4.4	206	1	RS4_SHIFL	P59132 shigella fl
45	59.5	4.4	206	1	RS4_YERPE	082j88 yersinia pe

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD; PRT; 191 AA.
ID YDA4_SCHPO
AC Q10346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F12.04c in chromosome I.
GN SPAC1F12.04c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
BX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holmes S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynolprez B.,
RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Punelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado B., Jimenez J., Sanchez M., del Rey F., Benito G.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Ussery D., Barrett B.G., Nurse P.,
RT The genome sequence of Schizosaccharomyces pombe.
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
DR EMBL; Z69944; CAAG3808.1; -
DR PIR; T38062; S67447
DR GeneDB_Spombe; SPAC1F12.04c; -

KW Hypothetical protein.
 SQ SEQUENCE 191 AA; 21549 MW; 65555347FOBBED16 CRC64;
 Query Match 5.8%; Score 78; DB 1; Length 191;
 Best Local Similarity 22.6%; Pred. No. 19;
 Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
 QY 104 MSDNTAANLLITTCGKRELTAFRLHNNGDVTRLDREPELNPAIPDERDTTPVAMAT 163
 1 MSYHSLNLMQNPBGIDKIALILVN---VARLD-----PASSKSTQVLSMLN 46
 DB 164 TSKRLTGLTLLASRQQLIDMMEADKVAPLRSAIPAGWFIADSGAERSGRTIA 223
 47 EPRC-----ILRLGLYKLLVNFRRKSSPEYMSMAINIGYVTE--GLAFICGKQITST 99
 QY 224 LGPDGKP-----SRIVYITTSQATMDERNROI----- 252
 100 -----SKLEDKMLMSSRFMLDTLTLTYQLREKTEDEKHOQLDASNLASLPICITHS 155
 QY 253 AETGASLIRK 262
 DB 156 VENGAGLHKH 165
 RESULT 2
 OGG1_ARCFU STANDARD; PRT; 198 AA.
 ID OGG1_ARCFU
 AC 029876;
 DT 16-OCT-2001 (Rel. 4.0, Created)
 DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
 DT 28-FEB-2003 (Rel. 4.1, Last annotation update)
 DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
 glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
 (EC 4.2.99.18) (AP lyase)].
 GN OGG OR AF0371.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 CC Archaeoglobaceae; Archaeoglobus.
 CX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
 Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
 Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
 Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 Maason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus.";
 RT Nature 390:364-370(1997).
 RL Nature 390:364-370(1997).
 CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
 guanine (7,8-dihydro-8-oxoguanine - 7-oxoG) from DNA. Also nicks
 DNA at apurinic/aprimidinic sites (AP sites) (By similarity).
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
 apyrimidinic site in DNA is broken by a beta-elimination reaction,
 leaving a 3'-terminal unsaturated sugar and a product with a
 terminal 5'-phosphate.
 CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: AE001079; AAB90876.1; -
 DR PIR: C69296; C69296.
 DR TIGR: AF0371; -
 DR HAMAP: MF_00241; -; 1.
 DR InterPro: IPR003265; Endo_3c.
 DR Pfam: PF00730; Hnh-GPD; 1.
 DR SMART: SM00478; ENDO3C; 1.
 KW Hydrolyase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosylase;
 KM Multifunctional enzyme; Complete proteome.
 FT ACT SITE 122 122
 FT SITE 122 122
 SQ SEQUENCE 198 AA; 22639 MW; 3A5C033A11F3FB CRC64;
 Query Match 5.3%; Score 71; DB 1; Length 198;
 Best Local Similarity 23.6%; Pred. No. 75;
 Matches 50; Conservative 26; Mismatches 80; Indels 56; Gaps 12;
 QY 14 QLGARVYIELDINSGEILSPREPERFPMSTFKVILGAVL---SRIDAG---QEQIG 67
 15 QLGEK-----GEVEPDFRPFLDFSVKATIRTELAFCTIANSATAGLKFQRLIG 64
 DB 68 RRIHYSQ-----NDLVEYSPYTERKHLTDGTVRELCSAITMSDNTAANLLITTI 117
 65 QGVGVKALTLAGVRFNRAEY-----IRAFKSFYKVEKALEAESKAREILLIK 117
 QY 118 G-GPKELTAFRLHNNG-DHYTRLD---RWPELNPAIPDERDTTPVAMATLTKLL-- 169
 DB 118 GLGKREASHFLRVNGRDVAIIDRHILRMEROGYEVPG---TWAKKYLEVERILME 172
 QY 170 -----TGLTLLASRQQLIDMMEADKVAPLR 197
 DB 173 ISERGESTL---AEMDLRIAE---MGKYLK 198
 RESULT 3
 RAC2_LOTJA STANDARD; PRT; 196 AA.
 ID RAC2_LOTJA
 AC 040220;
 DT 01-NOV-1997 (Rel. 3.5, Created)
 DT 01-NOV-1997 (Rel. 3.5, Last sequence update)
 DT 15-JUL-1998 (Rel. 3.6, Last annotation update)
 DE RAC2-Like GTP binding protein RAC2.
 GN RAC2.
 OS Lotus japonicus.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
 CX NCBI_TaxID=34305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Gifu / B-129; TISSUE=Root nodules;
 RX MEDLINE=97231679; PubMed=9076991;
 RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
 RT Identification of new protein species among 33 different small GTP-
 binding proteins encoded by cDNAs from Lotus japonicus, and
 RT expression of corresponding mRNAs in developing root nodules.";
 RL Plant J. 11:237-250(1997).
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

```

DR PRINTS: PR00449; RASTRNSFRMNG.
DR SMART: SM00174; RHO; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 13 20 GTP (BY SIMILARITY).
FT NP_BIND 60 64 GTP (BY SIMILARITY).
FT NP_BIND 118 121 GTP (BY SIMILARITY).
FT DOMAIN 35 43 EFFECTOR REGION (POTENTIAL).
FT LIPID 193 193 GERANYL-GERANYL (POTENTIAL).
SQ SEQUENCE 196 AA; 21704 MW; 8B80A70C4828C64; CRC64;

Query Match
Best Local Similarity 24.1%; Score 69.5; DB 1; Length 196;
Matches 45; Conservative 24; Mismatches 63; Indels 55; Gaps 11;

OY 44 MSTFKVLLCGAVLSRIDAGQEQGLRR--IHYSONDL-VETSPYTEKHLTDGMTVRELCSAA 99
DB 1 MSTAFKIC-----VTVGDGAVGKTCMLISYNTSPFDYVP-----TVFDNFS 44

OY 100 AATMSDNTANLLTTIGPKPE-----LTFAT-LHNGDHYTRLDRWEPE 143
DB 45 ANVVV-DQSTVNLGLMPTDAGQEDYRLRLPSYRGADVFLAFLSLSRASYENISKWIPPE 103

OY 144 LNEAIPNDERDTMPVAMATYLRKLTGELLTL-ASRQQLIDMEADKVA---GPLBSA 199
DB 104 LRHYAP-----TVPI-----VIVGTRKLDLRDQYLDHGCATPITTAGCEELKKA 149

OY 200 LPAGWPI 206
DB 150 IGAAYVL 156

RESULT 4
RACG_DICDI STANDARD; PRT; 201 AA.
AC QGPGSD; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ras-related protein racG.
GN Dictyostelium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=21127961; PubMed=11222756;
RA Rivero F., Dislich H., Glockner G., Noegel A.A.;
RT "The Dictyostelium discoideum family of Rho-related proteins.";
RL Nucleic Acids Res. 29:1068-1079(2001).
CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF10893; AAC45130.1; -.
DR HSSP: P21181; IAMA.
DR DictyDb; D077777; racG.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR001806; Ras_trnsfrmg.
DR Pfam; PR00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMNG.
DR SMART; SM00174; RHO; 1.
DR TIGRFAMS; TIGR00231; small_GTP; 1.
DR GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 10 17 GTP (BY SIMILARITY).

```

```

FT NP_BIND 57 61 GTP (BY SIMILARITY).
FT NP_BIND 115 118 GTP (BY SIMILARITY).
FT DOMAIN 32 40 EFFECTOR REGION (POTENTIAL).
FT LIPID 198 198 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 201 AA; 22813 MW; F6CE586497CC169 CRC64;

Query Match
Best Local Similarity 22.0%; Score 69; DB 1; Length 201;
Matches 35; Conservative 20; Mismatches 60; Indels 44; Gaps 7;

OY 44 MSTFKVLLCGAVLSRIDAGQEQGLRR-IHYSONDLV-EYSP-VTEKHLTDGMTVRELCSAA 101
DB 1 MRSIKVCVVG-----EGGIGKTSMLSTYSNSISNETPTVDNST----- 42

OY 102 ITMSDNTANALLTTIGPKELTAF-----LHNGDHYTRLDRWEPELN 145
DB 43 LHMHNKRPYNLSMDTAGQEEFSRLRLSPQDVFLLCPSLNPSSFSNILDSPQELN 102

OY 146 EAIPNDERDTMPVAMATYLRKLTGELL--TLASRQOL 182
DB 103 ENCPN-----TPIVLVGTQMDLKSNSVILDRICEKKOL 135

RESULT 5
HC3L_THIFE STANDARD; PRT; 122 AA.
AC P80509;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
DE (Fragments).
GN HOXG.
OS Rhodococcus ferrooxidans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
OX Acidithiobacillaceae; Acidithiobacillus.
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 19859;
RX MEDLINE=96241862; PubMed=8661919;
RA Fischer J., Quentmeier A., Kostka S., Kraft R., Friedrich C.G.;
RT "Purification and characterization of the hydrogenase from
RT Rhodococcus ferrooxidans.";
RL Arch. Microbiol. 165:289-296(1996).
CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferri-cytochrome c3 + 2
CC ferri-cytochrome c3.
CC -1- COFACTOR: IRON.
CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
CC -----
KW Oxidoreductase; Iron.
FT NP_BIND 20 21
FT NP_BIND 29 30
FT NP_BIND 35 36
FT NP_BIND 42 43
FT NP_BIND 59 60
FT NP_BIND 72 73
FT NP_BIND 78 79
FT NP_BIND 87 88
FT NP_BIND 98 99
FT NP_BIND 107 108
FT NP_BIND 122 122
SQ SEQUENCE 122 AA; 13146 MW; EB4FAF365348DA0B CRC64;

Query Match
Best Local Similarity 5.0%; Score 67; DB 1; Length 122;
Matches 28; Conservative 13; Mismatches 23; Indels 42; Gaps 4;

OY 19 VYIETDLSGFIILSPREERFPYMSYFKVLLCGAVLSRIDAGQEQGLRHYSONDLV 78
DB 12 VERVSGDLDSSILTYFR-----NALIARCGG-----LG 41

OY 79 EYSPYTEKHLTDGMTVRELCSAAITMSDNTANALLTTIGPKELT 124

```

Db 42 KTAAPGTGTEIGVT-----ISGDKDPQAGLVVT---PREST 75

RESULT 6

NDK_TREPA STANDARD; PRT: 149 AA.

AC 083974; 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)

DE (Nucleoside 2'-p kinase).

GN NDK OR TP1010.

OS Treponema pallidum.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI_TaxID=160;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Nichols;

RX MEDLINE=96832770; PubMed=9665876;

RA Fraser C.M., Norris S.J., Welnslock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Childsbarrow M., Utecherback T.,

RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of Treponema pallidum, the syphilis

RT spirochete.";

RL Science 281:375-388(1998).

CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES

CC OTHER THAN ATP.

CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate -> ADP +

CC nucleoside triphosphate.

CC -1- SUBUNIT: Homotrimer (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).

CC -1- SIMILARITY: Belongs to the NDK family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AE001269; AAC65961.1; -

DR PIR: F71252; F71252.

DR HSSP: P22887; INPK.

DR TIGR: TP1010; -

DR HAMAP: MF_00451; -; 1.

DR InterPro: IPR001564; NDK.

DR Pfam: PF00334; NDK; 1.

DR PRINTS: PR01243; NUCDPKINASE.

DR ProDom: PD001018; NDK; 1.

DR SMART: SM00562; NDK; 1.

DR PROSITE: PS00469; NDP_KINASES; 1.

KW Transferase; Kinase; ATP-binding; Complete proteome.

FT ACT_SITE 117 117 BY SIMILARITY.

FT SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

Query Match 4.9%; Score 66; DB 1; Length 149;

Best Local Similarity 21.6%; Pred. No. 1.3e+02;

Matches 35; Conservative 24; Mismatches 65; Indels 38; Gaps 7;

Db 29 GELLESFRRPBERPMMSTFKVLICGAVLSRIDAGQQLRRRIHYSONDLVEYSPTVERL 88

11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1 11:1

21 GEVLRSRF---ERKGLVLTALRLC-----VDATREL---HYAHERKPPYSLIAYI 67

89 TQGMTRVELCSAINTSDMTAANLLLTIGGPRELFAFLHNGDHTYRLDRNPELENAI 148

68 TSAAPV-----ALAFKGENAISLVRTLCS-----TRVEHAQP---GTI 103

QY 149 PND-ERDTPMVAATTLRLKLTGLLTLASRQQLIDMWED 189

Db 104 RGDFALRTTIVIVASDPSAARELALYFSADQFVEMRGN 145

RESULT 7

RBFA_UREPA STANDARD; PRT: 116 AA.

AC Q9P0H0; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ribosome-binding factor A.

GN RBFA OR UD321.

OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1)

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.

OX NCBI_TaxID=134821;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Serovar 3;

RX MEDLINE=20500219; PubMed=11048724;

RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

RA Caesell G.H.;

RT "The complete sequence of the mucosal pathogen ureaplasma

RT urealyticum.";

RL Nature 407:757-762(2000).

CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not

CC with 30S subunits that are part of 70S ribosomes or polysomes).

CC Essential for efficient processing of 16S rRNA. May interact with

CC the 5'terminal helix region of 16S rRNA (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).

CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AE002129; AAF30730.1; -

DR HAMAP: MF_00003; -; 1.

DR InterPro: IPR000238; Rib_bind_factA.

DR Pfam: PF02033; RBFA; 1.

DR ProDom: PD007327; Rib_bind_factA; 1.

DR TIGRfams: TIGR00082; rbfA; 1.

DR PROSITE: PS01319; RBFA; FALSE_NEG.

KW rRNA processing; Complete proteome.

FT SEQUENCE 116 AA; 13247 MW; A79DCG71F0547514 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 116;

Best Local Similarity 28.9%; Pred. No. 1.1e+02;

Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

Db 144 INEAIPNDERDTMPVANAATTLRLKLTGEL-----LTLSRQQLIDMW-ADRYAGPLL 196

18 INNALANEINDIKIALIARYAVR--LSNDLSVAKIFLDHAKRESMLTKLEWNVNKS-LL 74

QY 197 RSALPAGV-----FIADKS 210

Db 75 RSKLAEMVTSYKVPRLRFVDET 97

RESULT 8

KTHY_METAC STANDARD; PRT: 206 AA.

AC Q8THS9; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).

GN TMR OR M44433.
 OS Methanosarcina acetivorans.
 CC Archaea; Euryarchaeota; orders incertae sedis;
 CC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., DeRellano K., Johnson R.,
 RA Linton L., McEwen P., McKernan K., Talmas J., Tirlrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grisham D.A., Guss A.M.,
 RA Hedderich A., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Zander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate - ADP + thymidine
 CC 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE011164; AM07774.1; -
 DR HAMAP: MF_00165; -1-
 DR InterPro: IPR000062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1.
 DR TIGRPFAM: TIGR00041; DTMK_kinase; 1.
 DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
 DR Transference: Kinase; Nucleotide biosynthesis; ATP-binding;
 KW Complete proteome.
 KM NE_BIND 10 17 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 23303 MW; A19C25B5C85423B3 CRC64;
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.2e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 QY 56 LSRID-AGGEOGRRIHYSQNDIVESPTEKHLTDGKTVRELCASATMSDNTANLLI 114
 DB 8 LERIDSSGKSTVAKKIQ-ENSELDVPEVFTREPTRG-----TLTGNAVENAIO 55
 QY 115 TTIGGKRELTALHNHGDVHTRLDWEPELNEAIPDEBDTTPVAMATLRLKLGELL 174
 DB 56 SDPDQALFLFTADAEHLAKLVKPALEDGKTVISDSRYANOGITLKRNDPL- 114
 QY 175 TLASRQQLDWEADRVAGPLRLSALPAGW-----FLAD-----KSG-AGE----- 214
 DB 115 -----EWVR-----DLHKGVTIYPLDLTFLEDEPELAVNRCKRGSGQTF 154
 QY 215 -----RGSRGITIALGPDGKPSRIVYITTSQATMDE 247
 DB 155 EKIEPLRGVRELFLGLAIE-EPERFIVDASGSPEDVER 192
 RESULT 9
 ALKH_BACSU STANDARD; PRT; 196 AA.
 AC P50846;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE KMG/KDPG aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase
 DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KMG-aldolase); 2-
 DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
 DE dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate
 DE aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
 DE aldolase)].
 GN KOGA
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 RA Serror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kds loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.U., Conerton I.F., Cummings N.J., Carter N.M.,
 RA Danhot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enlart K.D., Errington J., Fabret C., Ferrari E., Poulger D.,
 RA Feltz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Colightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Hage K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Iaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koehler P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Maesuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Potwolik S., Prescott A.M.,
 RA Presecan E., Puig C., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffier F.,
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeruchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Wellenzeger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate - pyruvate +
 CC glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate -
 CC pyruvate + D-glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DODSONOFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GLYOXYLATE.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KMG/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: LA7838; AB38480.1; -.
DR EMBL: Z99115; CAB1427.1; -.
DR PIR: H69647; H69647.
DR Subtilisin; Bg11396; kdaA.
DR InterPro: IPR000887; Aldase_KDPG_KHG.
DR Pfam: PF01081; Aldolase; 1.
DR TIGRFAMs: TIGR01182; eda; 1.
DR PROSITE: PS00159; ALDOLASE_KDPG_KHG_1; FALSE_NEG.
DR PROSITE: PS00160; ALDOLASE_KDPG_KHG_2; 1.
DR Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
FT ACT_SITE 43 43 BY SIMILARITY.
FT ACT_SITE 47 47 BY SIMILARITY.
FT BINDING 130 130 SCHIFF-BASE WITH KHG OR PYRUVATE (BY
SEQUENCE 196 AA; 20865 MW; EB0B597FD1102CC4 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 196;
Best Local Similarity 21.6%; Pred. No. 3e+02;
Matches 27; Conservative 18; Mismatches 61; Indels 19; Gaps 4;

OY 28 SGELESPREPERPMSTFVLLCGAVLSRIDAGQ-EQLGRRIHYQN--DLVEYSPV 83
DB 51 ASDIEESFRNEDI-----LIGAGTVISAQAGAKAGAPVSPGSADLAELHSF 103
OY 84 TEKHLTGDMTVRELCSNAITMSDNTAANL-----LTTIGGPKELTAFLHMGDHVT 135
DB 104 VKTHYIGVLTPESELMALTFGFTTLKFPSCVGIPIFMKMLADPFGVTFPIPGIHP 163
OY 136 RLDRW 140
DB 164 EVDPM 168

RESULT 10
ID R4_PELINE STANDARD; PRT; 200 AA.
AC Q9FSD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Pellia necasiana (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellineae;
OC Pellaceae; Pellia.
OX NCBI_TaxID=70144;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=gametophyte;
RA Capesius I., Blocher R.;
RT "A molecular approach to bryophyte systematics.";
RU Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -!- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -!- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -!- SIMILARITY: BELONGS TO THE SAP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ250456; CA014050.1; -.
DR HSSP: P81288; 1C05.
DR HAMAP: MF_01306; -. 1.
DR InterPro: IPR001912; Ribosomal_S4.
DR InterPro: IPR002942; S4.
DR InterPro: IPR005709; S4_bact.org.
DR Pfam: PF00163; Ribosomal_S4; 1.
DR Pfam: PF01479; S4; 1.
DR SMART: SM00363; S4; 1.
DR TIGRFAMs: TIGR01017; rpsD_bact; 1.
DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
DR PROSITE: PS00889; S4; 1.
DR Ribosomal protein; RNA-binding; Chloroplast.
FT DOMAIN 90 150 S4 RNA-BINDING.
FT SEQUENCE 200 AA; 23178 MW; 00D00986D9333F5 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 200;
Best Local Similarity 23.4%; Pred. No. 3.1e+02;
Matches 40; Conservative 30; Mismatches 78; Indels 23; Gaps 8;

OY 14 QLGARVYIE--LDNNGELIESPREPERPMSTFVLLCGAVLSRIDAGQELGRRIH 71
DB 14 RLGLPGLTSLKLEIESGYIGOS--TPNKK--VSQYIRL-----EEKQKRFH 58
OY 72 Y--SQNDLVESPYTERKHLTDGMTRELCSNAITMSDNTAANL-LTTIGGPKELTAFLH 128
DB 59 YGLTERLLKRVARAR--AKSGTGLISQLEMLRNLNITFRGLMSPTIGARGLVNHRH 116
OY 129 NM-GDHYRLDRWEPELNEALPNDERDTTPVANATTLKILIGELLITLAS 178
DB 117 ILINDNTVDIPSYNCEPRDVTYNNRRESVYIKKMDSSRRKPKVPHNLTFS 167

RESULT 11
ID HSLV_THEME STANDARD; PRT; 176 AA.
AC Q9WY21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease hslV (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Hatt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher C.A., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RU Nature 399:323-329(1999).
CC -!- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -!- SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AF001728; AAD35606.1; -.
DR PIR: G72365; G72365.
DR HSSP: P31059; 1E94.
DR MEROPS: T01.006; -.
DR TIGR: TM0521; -.
DR HAMAP: MF_00248; -. 1.
DR InterPro: IPR001353; Proteome-protease.
DR Pfam: PF00227; Proteasome; 1.
DR Hydrolase: Protease; Complete proteome.
KW ACT_SITE 6
SQ SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 2.9e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 IGGPKETAPLHMNDHVTLDKREPELNAIPNDRDPTMPPVAMATTLKLTGLTL 176
DB 42 LGSGKVLVAGPAGSVADAMTLEDFEAKLR-----WGNTLR 78
QY 177 ASRQQLIDMNEADKVAGPLRSALPAGWFIADR-----SGAGERSGII-----AAL 224
DB 79 AAVEAKDW-RTDRV-----LR-RLBALLLVADKENFIISNGE-----VIOPDDAAI 127
QY 225 GPDC 228
DB 128 GSGG 131

RESULT 12
MARR_ECOLI STANDARD; PRT; 144 AA.
AC P27245; P76882; P77582;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Multiple antibiotic resistance protein marr.
GN MARR OR SOX OR CFXB OR INAR OR B1530.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186717; PubMed=8383113;
RT Cohen S.P., Haechler H., Levy S.B.;
RT Genetic and functional analysis of the multiple antibiotic
RT resistance (mar) locus in Escherichia coli.
RT J. Bacteriol. 175:1484-1492(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA *The complete genome sequence of Escherichia coli K-12.
RT Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Kasai H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itch T.,
RA Makino K., Maki T., Mizobuchi K., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Oshima T., Saito N.,
RA Samped G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT *A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.
RT DNA Res. 3:363-377(1996).

```

```

RN [4]
RP CHARACTERIZATION.
RX MEDLINE=94110219; PubMed=8282690;
RA Ariza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;
RT *Repressor mutations in the marRAB operon that activate oxidative
RT stress genes and multiple antibiotic resistance in Escherichia
RT coli.
RT J. Bacteriol. 176:143-148(1994).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=95286534; PubMed=7768850;
RA Seoane A.S., Levy S.B.;
RT *Characterization of MarR, the repressor of the multiple antibiotic
RT resistance (mar) operon in Escherichia coli.
RT J. Bacteriol. 177:3414-3419(1995).
RN [6]
RP MUTAGENESIS.
RX MEDLINE=20223625; PubMed=10760140;
RA Alekshun M.N., Kim Y.S., Levy S.B.;
RT *Mutational analysis of MarR, the negative regulator of marRAB
RT expression in Escherichia coli, suggests the presence of two regions
RT required for DNA binding.
RT Mol. Microbiol. 35:1394-1404(2000).
CC -1- FUNCTION: REPRESSOR OF THE MARRAB OPERON WHICH IS INVOLVED IN THE
CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS
CC GENES. BINDS TO THE MARO OPERATOR/PROMOTER SITE.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC
DR EMBL: M96235; AAC16394.1; ALT_INIT.
DR EMBL: AE000250; AAC74603.1; ALT_INIT.
DR EMBL: D90795; BAA15212.1; -.
DR EMBL: D90796; BAA15220.1; -.
DR EMBL: D90797; BAA15232.1; -.
DR PIR: E64907; E64907.
DR PDB: 1UGS; 28-DEC-01.
DR Ecogene; Egl1435; marr.
DR InterPro: IPR000835; HTH_Marr.
DR Pfam: PF01047; HTH_Marr.
DR PRINTS: PR00598; HTHMARR.
DR SMART: SM00347; HTH_MARR.1.
DR PROSITE: PS01117; HTH_MARR_FAMILY.1.
KW TRANSCRIPTION REGULATION; DNA-binding; Repressor;
KW Antibiotic resistance; Complete proteome; 3D-structure.
FT MUTAGEN 45
FT MUTAGEN 45
FT MUTAGEN 77
FT MUTAGEN 123
FT MUTAGEN 144
FT SEQUENCE 144 AA; 16065 MW; BE7DF5549E24D1D3 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 144;
Best Local Similarity 29.4%; Pred. No. 2.4e+02;
Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

QY 66 LGRRIHY--SQND--LVRY--SP--VTEKHLIDGMVRELCSAATMSNTAANLLTIG 118
DB 14 LGRLLHMNQKRLNLESLPDLTTAAQFVLCISR--CAACT----- 56
QY 119 GPKETAPLH-NMGDHTVRLDR-----WEPELNEAIPNDRDPTMPPVAMATTLKLTG- 171
DB 57 -PVELKVLVSDIGALTRLDRLVCKGVERLPN--PDKNG-----VLVKLTGG 104
QY 172 -----ELTLASRQQLIDMNEADKVA--GPLRSALP 201

```

Db 105 AAICECHOLVGODLHQELTKNLTADEVALEYLLKKVLP 144

RESULT 13

YDEJ_ECOLI STANDARD; PRT; 172 AA.
 AC P3131;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein ydeJ.
 GN YDEJ OR B1537.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 CX NCBI_TaxID=562;

SEQUENCE FROM N.A.
 RX MEDLINE=93186717; PubMed=8383113;
 RA Cohen S.P., Haeckler H., Levy S.B.;
 RT "Genetic and functional analysis of the multiple antibiotic
 resistance (mar) locus in Escherichia coli.";
 RL J. Bacteriol. 175:1484-1492(1993).

SEQUENCE FROM N.A.
 RX MEDLINE=97426517; PubMed=9278503;
 RA Blittner F.R., Plunkett G., Ilt, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Grogan J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).

SEQUENCE FROM N.A.
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alia H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 Itoh T., Kasai H., Kasahiro K., Kimura S., Kitakawa M.,
 Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,
 Tagami H., Takeda J., Takeuchi K., Takeuchi Y., Wada C.,
 Yamamoto Y., Horikuchi T.;
 RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).

CC -1- SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E.COLI YGAD.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; M96235; -; NOT ANNOTATED_CDS.
 DR EMBL; AE000251; AAC74610.1; -;
 DR EMBL; D90796; BAA15227.1; -;
 DR EMBL; D90797; BAA15239.1; -;
 DR PIR; D64908; D64908.
 DR ECOSGENE; EG11645; ydeJ.
 DR HAMAP; MF_00236; atypcJ. 1.
 DR InterPro; IPR003396; CINA.
 DR Pfam; PF02464; CINA; 1.
 DR TIGRFAMs; TIGR00199; cina.cterm; 1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 172 AA; 18321 MW; 700A7A95C967C868 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 172;
 Best Local Similarity 27.1%; Pred. No. 3.1e+02;

Matches 23; Conservative 9; Mismatches 26; Indels 27; Gaps 3;
 QY 73 SONDVEYSPTEKHLTDGVTRELCSPAITMSDNTANLLTTGSP----- 120
 Db 79 SOOLEYTSVSRK-----VAEMATGAIERAD-ADVSAITGYGGPEGEGCTPAGTV 131

QY 121 -----KELTAFLHMGDHYTRL 137
 Db 132 WFAWHKQNYTAVMHFAGDCETVL 156

RESULT 14

PYRE_STRCO STANDARD; PRT; 182 AA.
 ID PYRE_STRCO
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT) (OPRTase).
 GN PYRE OR SC03650 OR SCH10.28C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycinae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=1902;

SEQUENCE FROM N.A.
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
 Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate + diphosphate -> orotate
 CC + 5-phospho-alpha-D-ribose 1-diphosphate.
 CC -1- Cofactor: Magnesium (by similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; fifth step.
 CC -1- SUBUNIT: Homodimer (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY. PYRE SUBFAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC EMBL; AL939117; CAB42037.1; -;
 DR PIR; T36540; T36540.
 DR HAMAP; MF_01208; -; 1.
 DR InterPro; IPR004467; Or-phospho.trans.
 DR InterPro; IPR002375; pr/py-rp.transf.
 DR InterPro; IPR000836; prtransf.
 DR Pfam; PF00156; Pribosyltran; 1.
 DR TIGRFAMs; TIGR00336; pyre; 1.
 DR PROSITE; PS00103; PUR-PYR.PR.TRANSFER.FALSE_NEG.
 KW Pyrimidine biosynthesis; Transferase; Glycosyltransferase; Magnesium;
 KW Complete proteome.
 FT DOMAIN 118 130 PRPP-BINDING (BY SIMILARITY).
 FT ACT_SITE 100 100 BY SIMILARITY.
 SO SEQUENCE 182 AA; 19082 MW; FD5DA4AA3089137BC CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 182;
 Best Local Similarity 28.8%; Pred. No. 3.3e+02;

Matches 34; Conservative 14; Mismatches 41; Indels 29; Gaps 7;

```
OY 165 LRKL-LTGELLTLASRQQLD-----WMEADKVAQPLRSALPAG---WFI 206
Db 36 LRRTLDGEAPLVG-QVLLDLTLADLEFDVAVGGLTGMADPVAAASMLHAARGRLDAFV 94
OY 207 ADKSGAGRGSRGIIALGPGKPSRIYVITTT---GSQATMDERNQIAEIGASLI 260
Db 95 VRKTAKAHGLQRRV---EGPEIKGRRVVVEDTGTGSGPLTAVE---AVREACAEVY 146
```

RESULT 15

```
TRPE_THEVO
ID TRPE_THEVO STANDARD: PRT: 195 AA.
AC Q979V6;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).
GN TRPE OR TVI054 OR TVGI080245.
OS Thermoplasma volcanum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasmata.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima T., Yamamoto Y., Watanabe K., Yamazaki M., Kaneshiro K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanum."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC -1- CARBAMYLIC ACTIVITY: N-(5-phospho-beta-D-riboseyl)-anthranilate - 1-
CC (2-carboxyphenylamino)-1-deoxy-D-ribose 5-phosphate.
CC -1- PATHWAY: Tryptophan biosynthesis; third step.
CC -1- SIMILARITY: BELONGS TO THE TRPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000895; BAB60196.1; .
DR HAMAP: MF_00135; -; 1.
DR InterPro: IPR001240; PRAI.
DR Pfam: PF00697; PRAI; 1.
KM Isomerase: Tryptophan biosynthesis; Complete proteome.
SQ SEQUENCE 195 AA; 21325 MW; FE3C91205E080D3 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 195;
Best Local Similarity 27.0%; Pred. No. 3.6e+02;
Matches 51; Conservative 21; Mismatches 90; Indels 27; Gaps 9;

OY 48 KVLICGAVLSRIDAGQGLGRIRHYSONDLVEISPTVEKRLTDGKTVRELCSAATMSDN 107
Db 2 KIVVCGITRLEDAAMATELGASIVGVLD--ELSPRHASH---NTIREIAEGITVAGV 55
OY 108 TAANLLLTITIGPK-ELTAFLEHNMGDHVTRLDMEPELNEAIPNDEKDTMPVAAATTLR 166
Db 56 YTSE--QTVLSPLEFDYVQLH--FDH-----DPELIRSIHELGR-KVISYINNGIR 103
OY 167 KLLTGELTLASRQQLIDMEADKVAQPLRSALPAGWFIADKSGAGRGSRGIIALGP 226
Db 104 DL---KIKYNAYREADIILVEYKGVSVISQIAPLGLNMGYAGGISDRDIENITAA---- 157
OY 227 DCKPSRIIV 235
Db 158 --KPSIIDV 164
```

Search completed: September 10, 2003, 12:20:49
Job time : 7.57143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E

Sequence: 1 HPEPLVAVDAEDQIARVG.....TMDERNKQIAIGASLIKRW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	534	38.9	105	2	Q52026 pseudomonas
2	524	38.9	105	2	Q52330 escherichia
3	493	36.6	145	2	Q8RTD8 klebsiella
4	488	36.2	102	2	Q52639 pseudomonas
5	458	34.0	138	2	Q990F3 escherichia
6	458	34.0	138	2	Q9AMA1 escherichia
7	458	34.0	139	2	Q9AMA2 escherichia
8	456	33.8	138	2	Q9AMA0 escherichia
9	456	33.8	138	2	Q9AMA9 escherichia
10	456	33.8	139	2	Q9AM98 escherichia
11	453	33.6	139	2	Q9AMA3 escherichia
12	338	25.1	67	2	Q53553 shigella fl
13	320	23.7	95	2	Q8GDES shigella fl
14	316	23.4	62	2	Q9JN58 shigella fl
15	299	22.2	128	2	Q8VQ00 escherichia
16	246	18.2	52	2	Q9R412 shigella fl

17	245	18.2	180	2	Q8KVT2 staphylococ
18	241	17.9	48	2	Q9RIH0 proteus mir
19	197	14.6	38	2	P97145 escherichia
20	164.5	12.2	134	2	Q8V0L3 staphylococ
21	119.5	8.9	109	2	Q53698 staphylococ
22	105	7.8	20	2	P97146 escherichia
23	87	6.5	100	2	Q93S05 staphylococ
24	82	6.1	68	2	Q9XJ2 bacillus ce
25	79	5.9	198	2	Q9ACM8 streptococ
26	77	5.7	181	16	Q8D108 yeastina pe
27	75.5	5.6	190	5	076227 trypanosoma
28	75.5	5.6	200	16	Q9JUE8 clostridium
29	75.5	5.6	202	16	Q9S520 rhizobium l
30	74	5.5	177	16	Q9KEM7 bacillus ha
31	73.5	5.5	195	16	Q9RIH0 streptococ
32	73.5	5.5	204	11	Q9IYW4 mus musculu
33	73	5.4	153	5	Q8SS20 o8a28 pyrobaculum
34	73	5.4	175	17	Q8ZTH8 o0584 staphylococ
35	72.5	5.4	113	2	Q49970 o9191 streptococ
36	72.5	5.4	145	2	Q05984 o9x9h0 yeastina en
37	71.5	5.3	152	16	Q9R191 o972w2 sulfobius
38	71	5.3	131	2	Q9XSH0 o9hw07 pseudomonas
39	71	5.3	169	17	Q9XSH0 o9hw07 pseudomonas
40	71	5.3	192	16	Q9HM07 o9x9h0 yeastina en
41	70.5	5.3	203	4	Q96B23 o9hw07 pseudomonas
42	70.5	5.2	131	16	Q8XOL8 o9x9h0 yeastina en
43	70.5	5.2	184	5	Q45379 o9hw07 pseudomonas
44	70	5.2	108	17	Q9HR03 o9x9h0 yeastina en
45	70	5.2	150	10	Q94IG4 o9x9h0 yeastina en

ALIGNMENTS

RESULT 1	ID	Q52026	PRELIMINARY;	PRT;	105 AA.
AC	Q52026	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)			
DE	Beta-lactamase (Fragment).				
GN	BLA.				
OS	Pseudomonas aeruginosa.				
OG	Plasmid pR01614.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Pseudomonas.				
OX	NCBI_TaxID=287;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-95011664; PubMed-7926843;				
RA	West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;				
RT	*Construction of improved Escherichia-Pseudomonas shuttle vectors				
RT	derived from pUC18/19 and sequence of the region required for their				
RT	replication in Pseudomonas aeruginosa.;				
RU	Gene 148:81-86(1994).				
RU	Gene 148:81-86(1994).				
DR	EMBL; U30112; AAA66058.1; -.				
DR	HSSP; P00810; IXPB.				
DR	InterPro: IPR001466; Beta_lactamase.				
DR	InterPro: IPR000871; Beta_lactamase.A.				
DR	Pfam: PF00144; beta_lactamase; 1.				
DR	PRINTS; PR00118; BLACTAMASEA.				
KW	Plasmid.				
FT	NON_TER				
FT	SEQUENCE	105 AA; 11229 MW; D2889A4073330557 CRC64;			
Query Match		38.9%; Score 524; DB 2; Length 105;			
Best Local Similarity		100.0%; Pred. No. 1e-36;			
Matches	104; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	160	AMATTIRLRLNGELTLTASROOLDPMEDADYVAGPLRLSALPAGFTADKSGAGERSRG	219		
DB	2	AMATTIRLRLNGELTLTASROOLDPMEDADYVAGPLRLSALPAGFTADKSGAGERSRG	61		

```

OY      220 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 263
      |||||||
Db      62 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 105
      |||||||

RESULT 2
ID      052330      PRELIMINARY;      PRT;      105 AA.
AC      052330;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Tnl b1a protein (Fragment).
OS      Escherichia coli.
OC      Plasmid RK2.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90264294; PubMed=2160936;
RA      Kornacki J.A., Burlage R.S., Figurski D.H.;
RT      "The k11-for regulation of broad host-range plasmid RK2: Nucleotide
RT      sequence, polypeptide product and expression of regulatory gene
RT      korc";
RL      J. Bacteriol. 172:3040-3050(1990).
DR      EMBL: M32794; AAA26408.1; -.
DR      HSSP: P00810; 1XPB.
DR      InterPro: IPR001466; Beta_lactamase.
DR      InterPro: IPR000871; Beta_lactamase_A.
DR      Pfam: PF00144; beta_lactamase; 1.
DR      PRINTS: PR00118; BLACTAMASEA.
KW      Plasmid.
FT      NON_TER      1
SQ      SEQUENCE      105 AA;      11229 MW;      D28894407330557 CRC64;

Query Match      38.9%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 1e-36;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      160 AMATTIRKLITGELLTTLASRQQLIDWMEADRVAGPLRLSALPAGFTADKSGAGERSRG 219
      |||||||
Db      2 AMATTIRKLITGELLTTLASRQQLIDWMEADRVAGPLRLSALPAGFTADKSGAGERSRG 61
      |||||||

OY      220 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 263
      |||||||
Db      62 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 105
      |||||||

RESULT 3
OY      08RTD8      PRELIMINARY;      PRT;      145 AA.
AC      08RTD8;
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT      01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE      SHV-5 enzyme (Fragment).
OS      Klebsiella pneumoniae.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Klebsiella.
OX      NCBI_TaxID=573;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Palisudramaniam S.;
RT      "Characterization of SHV-5 extended-spectrum beta-lactamase from
RT      ceftriaxime-resistant Klebsiella pneumoniae";
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF467105; AAL75506.1; -.
DR      InterPro: IPR001466; Beta_lactamase.
DR      Pfam: PF00144; beta_lactamase; 1.
FT      NON_TER      1
FT      NON_TER      145

```

```

SQ      SEQUENCE      145 AA;      15574 MW;      F88634D6194BC82 CRC64;

Query Match      36.6%; Score 493; DB 2; Length 145;
Best Local Similarity 68.1%; Pred. No. 6.7e-34;
Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

OY      100 AATMSDNTAANLLTTTGPKELTAFLHNMGDVTRLRMEPELNEAIPNDERDTAPV 159
      |||||||
Db      7 AATYMSDMSANLLTATVGPGATLTAFLNOIGDVTLRDRWTELNEALPGDARDTTPA 66
      |||||||

OY      160 AMATTIRKLITGELLTTLASRQQLIDWMEADRVAGPLRLSALPAGFTADKSGAGERSRG 219
      |||||||
Db      67 SMAATIRKLITSORLSARSGRQLQWMDRVAGPLIRSVLPAGMFIADRTGASKRGARG 126
      |||||||

OY      220 IIALGPDGKPSRIVYIT 237
      |||||||
Db      127 IVALGPNNAERIVYIT 144
      |||||||

RESULT 4
ID      052639      PRELIMINARY;      PRT;      102 AA.
AC      052639;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      Beta-lactamase (Fragment).
OS      BLA.
OS      Pseudomonas aeruginosa.
OC      Plasmid PRO1600.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC      Pseudomonadaceae; Pseudomonas.
OX      NCBI_TaxID=287;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      TRANSPOSON-Tn1.
RA      MEDLINE=94336737; PubMed=8058819;
RA      Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
RT      "Deletion and transposon mutagenesis and sequence analysis of the pOR
RT      PRO1600 OriL region found in the broad-host-range plasmids of the pOR
RT      series";
RL      Plasmid 31:265-274(1994).
DR      EMBL: L22691; AAA98312.1; -.
DR      HSSP: P00810; 1XPB.
DR      InterPro: IPR001466; Beta_lactamase.
DR      InterPro: IPR000871; Beta_lactamase_A.
DR      Pfam: PF00144; beta_lactamase; 1.
DR      PRINTS: PR00118; BLACTAMASEA.
KW      Plasmid.
FT      NON_TER      1
SQ      SEQUENCE      102 AA;      10827 MW;      DA95AF7557DA13D4 CRC64;

Query Match      36.2%; Score 488; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.1e-33;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      160 AMATTIRKLITGELLTTLASRQQLIDWMEADRVAGPLRLSALPAGFTADKSGAGERSRG 219
      |||||||
Db      2 AMATTIRKLITGELLTTLASRQQLIDWMEADRVAGPLRLSALPAGFTADKSGAGERSRG 61
      |||||||

OY      220 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 257
      |||||||
Db      62 IIALGPDGKPSRIVYITGSOATMDERNROIAEIGASLTKM 99
      |||||||

RESULT 5
ID      0990F3      PRELIMINARY;      PRT;      138 AA.
AC      0990F3;
DT      01-JUN-2001 (TREMBLrel. 17, Created)
DT      01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)

```



```

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia.
DR EMBL: AF327049; AK07464.1; -
DR EMBL: AF327051; AK07466.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

```

```

Query Match 34.0%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 5.6e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFIHNNGDHYTRLDRWEPELNEALPNDERDTPVMAATTLRLLT 170
DB 10 NLLATVGGPAGLTAFIROIIGDVTRLDRWETELNEALPGDARDTTPASMAATTLRLLT 69
QY 171 GELLTLASROQLIDMNEADRVAGPLIRSLPAGMFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGMFIADKSGASKRGARIVALLGPNKA 129
QY 231 SRIVVY 237
DB 130 ERIVVY 136

```

```

RESULT 6
Q9AMA1 PRELIMINARY; PRT; 138 AA.
AC Q9AMA1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1B/25;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia.
DR EMBL: AF327050; AK07465.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.0%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 5.6e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

```

```

QY 111 NLLTTTGGPKELTAFIHNNGDHYTRLDRWEPELNEALPNDERDTPVMAATTLRLLT 170
DB 10 NLLATVGGPAGLTAFIROIIGDVTRLDRWETELNEALPGDARDTTPASMAATTLRLLT 69
QY 171 GELLTLASROQLIDMNEADRVAGPLIRSLPAGMFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGMFIADKSGASKRGARIVALLGPNKA 129
QY 231 SRIVVY 237
DB 130 ERIVVY 136

```

```

RESULT 7
Q9AMA2 PRELIMINARY; PRT; 139 AA.
AC Q9AMA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=918;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia.
DR EMBL: AF327048; AK07463.1; -
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

```

```

Query Match 34.0%; Score 458; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 5.7e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFIHNNGDHYTRLDRWEPELNEALPNDERDTPVMAATTLRLLT 170
DB 10 NLLATVGGPAGLTAFIROIIGDVTRLDRWETELNEALPGDARDTTPASMAATTLRLLT 69
QY 171 GELLTLASROQLIDMNEADRVAGPLIRSLPAGMFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGMFIADKSGASKRGARIVALLGPNKA 129
QY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 8
Q9AMA0 PRELIMINARY; PRT; 138 AA.
AC Q9AMA0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

RC	STRAIN-E/98 9-1;		
RA	Subramaniam G.; Navaratnam P;		
RT	SHV gene sequences from ceftriaxime-resistant <i>Escherichia coli</i>		
RL	isolates from UMCC, Malaysia.;		
DR	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF227052; AAK07467.1; -		
DR	HSP; P14557; ISHV.		
DR	Interpro: IPR001466; Beta_lactamase.		
DR	Interpro: IPR000871; Beta_lactamase_A.		
DR	Pfam: PF00144; beta_lactamase; 1.		
DR	PRINTS; PR00118; BLACTAMASEA.		
FT	NON_TER	1	
FT	NON_TER	138	
Q0	SEQUENCE	138 AA; 15105 MW; 561D0922F5442847 CRC64;	

Query Match	33.8%	Score	455	DB 21	Length	138	
Best Local Similarity	66.9%	Pred. No.	8.3e-31				
Matches	85	Conservative	24	Indels	0	Gaps	0

```

QY 111 NLLLTITGCKELTAFLLHMGDHVTRLDMEFLNEALPDERBRTTYPVAMATTLRLKLT 170
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 10 NLLLATVGGPAGTAFRLRGDNVTRLDLRMETELNEALPGDARRTTTPASMAATTLRLKLT 69
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 171 GELLTLASRQQLIDMHEADKAVGAPLLRSALPAGWTIADKSAGERSGIIAALGPCKP 230
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 70 SORLSASRQQLIDMHEADKAVGAPLLRSALPAGWTIADKSAGERSGIIAALGPCKP 129
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 231 SRTIVTY 237
    |||||:|
Db 130 ERTIVLY 136

```

RESULT 9

Q9AM99 PRELIMINARY; PRT; 138 AA.

DT 01-JUN-2001 (TREMblrel_17, Created)
DT 01-JUN-2001 (TREMblrel_17, Last sequence update)
DT 01-MAR-2002 (TREMblrel_20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E/99 4-1;
RA Subramaniam G., Navaratnam P.;
RT "SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT Isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF327053; AA07468.1; -.
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match	33.8%	Score 456	DB 2	Length 138
Best Local Similarity	66.9%	Pred. NO. 8.3e-31		
Matches 85; Conservative	18;	Mismatches 24;	Indels 0;	Gaps 0;

[illegible]

QY	231	SRIVVIX	237
		:	
Db	130	ERIVVLY	136

RESULT 10
Q9AM98

ID	Q9AM98	PRELIMINARY;	PRT;	139	AA.
----	--------	--------------	------	-----	-----

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E/99 3-2;
RA Sudranam Iam G., Navaratnam P.;
RT "SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327054; AA07469.1; -.
DR HSSP; P14557; ISHV.
DR InterPro; IPR001466; Beta_lactamase.
DR InterPro; IPR000871; Beta_lactamase_A.
DR Pfam; PF00144; beta_lactamase; 1.
DR PFam; PF00118; BLACRMASE.
FT NON_TER 1
FT TER 1
SQ SEQUENCE 139 AA; 15234 MW; 03361A792F5442A8 CRC64;

Query Match	33.8%	Score 456	DB 2	Length 139
Best Local Similarity	66.9%	Pred. No. 8.4e-31		
Matches 85; Conservative	18;	Mismatches 24;	Indels 0;	Gaps 0;

QY	111	NILLTIGGPKETLAF	LHMKGCHVFL	LDNREPLNEA	IPDEODITPM	PVAMATTLKILT	170
Db	10	NILLTAVGSPAGLTFAL	ROIGDVTFLD	NETELNEAL	IPDADDTTP	PASMAATTLRLLT	69
QY	171	GELLTILASGOQLD	MEADKRVAGP	LLRSALPACMF	LADSGAGERBS	RIIATLGPDKP	230
		: :	: :	: :	: :	: :	
		: :	: :	: :	: :	: :	
Db	70	SQRLSARSOROL	LOMNVDDRVAGP	TLRSVLPACMF	LADTKGASK	RGARCIALLGPNNKA	129
QY							
	231	SRIIVLY	237				
Db			130	ERIVLY	136		

RESULT 11

Q9AMA3 PRELIMINARY; PRT; 139 AA

DT 01-JUN-2001 (TREMBLER, 17, Created)
DT 01-JUN-2001 (TREMBLER, 17, Last sequence update)
DT 01-MAR-2002 (TREMBLER, 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:935;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from ceftriaxide-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327047; AAC07462.1; -.
DR HSSP; P14557; ISHV.
DR InterPro; IPR001466; Beta_lactamase.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds

(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E_157T

Perfect score: 1348

Sequence: 1 HPEHTLVKVADEQDLGARVG.....TMDENRQIAEIGASLIKRW 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

865366

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	43.5	196	7	AAp60627
2	571	42.4	146	22	AAU23221
3	571	42.4	146	22	AAU23221
4	492	36.5	127	18	AAU20440
5	477	35.4	129	22	AAU23220
6	477	35.4	129	22	AAU23220
7	477	35.4	129	22	AAU23220
8	457	33.9	94	21	AAU23220
9	404	30.0	159	7	AAU23220

10	366	27.2	101	22	ABG27935
11	333	24.7	88	22	ABG27919
12	199	14.8	119	7	AAp60626
13	141	10.5	51	21	AAU2783
14	80.5	6.0	158	22	AAU45138
15	78	5.8	15	10	AAU98503
16	77	5.7	200	18	AAU55524
17	75.5	5.6	170	19	AAU38722
18	75	5.6	202	24	AAU32763
19	75	5.6	202	24	AAU32763
20	71.5	5.3	196	21	AAU16283
21	71	5.3	159	19	AAU69170
22	70.5	5.2	193	23	AAU76854
23	70.5	5.2	193	23	AAU76854
24	70	5.2	170	22	AAU52472
25	70	5.2	202	22	AAU82686
26	69.5	5.2	170	22	AAU56330
27	69.5	5.2	184	21	AAU25446
28	69	5.1	134	21	AAU27220
29	69	5.1	166	23	ABP26537
30	69	5.1	194	22	AAU81764
31	68	5.0	180	21	AAU20206
32	67.5	5.0	152	22	ABG19138
33	67.5	5.0	187	21	AAU27771
34	67.5	5.0	187	21	AAU27771
35	67.5	5.0	187	21	AAU27771
36	67.5	5.0	187	21	AAU27771
37	67	5.0	187	21	AAU27771
38	67	5.0	187	21	AAU27771
39	66.5	4.9	183	22	AAU59092
40	66	4.9	183	22	AAU59092
41	66	4.9	202	24	AAU59092
42	65.5	4.9	50	17	AAU32764
43	65.5	4.9	102	23	AAU32764
44	65.5	4.9	136	21	AAU32764
45	65.5	4.9	177	21	AAU32764

ALIGNMENTS

RESULT 1	AAp60627	standard; Protein; 196 AA.
ID	AAp60627	standard; Protein; 196 AA.
XX	AAp60627	
AC	25-MAR-2003	(updated)
DT	17-JUN-1991	(first entry)
XX		
DE	Beta-urogastrone - beta-lactamase fusion protein from pUG2101.	
XX		
KW	Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds; fusion protein; beta-lactamase.	
KW		
XX		
OS	Synthetic.	
EH	Key	Location/Qualifiers
FT	Protein	1..120
FT	Misc-difference	121..123
FT	Protein	/label= adaptor
FT		/label= beta-urogastrone
XX		
XX	DE3523634-A.	
PN	09-JAN-1986.	
PD	02-JUL-1985;	85DE-3523634.
PF	02-JUL-1985;	85DE-3523634.
XX		
PR	02-JUL-1984;	84JP-0137691.
XX		

Novel human diago
Novel human diago
Beta-urogastrone -
Vtg98-beta-lactama
Protonibacterium
Sequence encoded b
H. pylori ORF 29ep
S. pneumoniae SGH
Human zcyto24 prot
Mouse IMX129840-2
N-termally tagge
Human integrin alp
Human integrin alp
Propionibacterium
S. epidermidis ope
Propionibacterium
Pinus radiata cell
Zea mays protein f
Streptococcus poly
S. epidermidis ope
Arabidopsis thalia
Novel human diago
Novel human diago
Arabidopsis thalia
Arabidopsis thalia
Human MK61 protein
C glutamicum prote
Streptococcus pneu
Propionibacterium
S. pneumoniae deri
Human zcyto25 prot
Human alpha 2 C4 e
Human dehydrogenas
Partial corn extra
Human ORF1346

PA (EART) EARTH CHEM CO LTD.
XX
PI AOKI S, Ohgaki H, Horinaka A, Hiramatsu H, Kounoto S, Nishimura A;
PI Matsushiro S;
XX WPI; 1986-015031/03.
DR N-PSDB; AAN60631.
XX
PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PI and transformed cells contg. it.
XX
PS Disclosure: Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
XX
SQ Sequence 196 AA:
Query Match 43.5%; Score 587; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. NO. 4,9e-53;
Matches 115; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 HPEITLVKVKDAEDDVGARVGIETLDLNGSEILIESFRPERPPMSTFVLLCGAVLSRID 60
DB 24 HPEITLVKVKDAEDDVGARVGIETLDLNGSKILESFRPERPPMSTFVLLCGAVLSRVD 83
OY 61 AGOEOLGRRIHSQNDLVEYSPVTEKHLTDGNTVRELCSAATMSDNTAANILLTTI 117
DB 84 AGOEOLGRRIHSQNDLVEYSPVTEKHLTDGNTVRELCSAATMSDNTAANILLTTI 140
RESULT 2
AAU23221
ID AAU23221 standard; Protein; 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200153301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233653.
PR 14-SEP-2000; 2000US-0233654.
PR 14-SEP-2000; 2000US-0233655.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

CC	diagnosis/treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. arteriosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy.
CC	AaU22915-AaU23814 represent the novel human enzyme polypeptides of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
CC	
CC	
xx	Sequence 146 AA;
SQ	
Query Match	42.4%; Score 571; DB 22; Length 146;
Best Local Similarity	97.4%; Pred. No. 1.5e-51;
Matches 112; Conservative	1; Mismatches 2; Indels 0; Gaps 0
OY	1 HPELVKKYKDAEDQLGANGVETLDLNSGEILSEFRPERFPMSTFKVLGCAYLSRID 60
Dd	28 HPETLVKKYKDAEDQLGANGVETLDLNSGKILSEFRPERFPMSTFKVLGCAYLSRID 87
OY	61 AGOELGRRIHYSNDLVEYSPYRKHLTDGNFTVELCSAATMSDNTANILLT 115
Dd	88 AGOELGRRIHSNDLVEYSPYRKHLTDGNFTVELCSAATMSDNTANILLT 142
RESULT 3	
ID	AAM90870
AA	AAM90870 standard; Protein; 146 AA.
xx	
AC	AAM90870;
xx	
DT	07-NOV-2001 (first entry)
xx	
DE	Human Immune/Haematopoietic antigen SEQ ID NO:18463.
xx	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KX	Cytostatic; gene therapy; vaccine; metastasis.
OS	Homo sapiens.
xx	
PN	MO200157182-A2.
PD	
xx	
xx	09-AUG-2001.
PF	17-JAN-2001; 2001WO-US01354.
xx	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184666.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209465.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.

PR	14-AUG-2000	200005-0225266
PR	14-AUG-2000	200005-0225267
PR	14-AUG-2000	200005-0225268
PR	14-AUG-2000	200005-0225270
PR	14-AUG-2000	200005-0225447
PR	14-AUG-2000	200005-0225575
PR	14-AUG-2000	200005-0225758
PR	14-AUG-2000	200005-0225759
PR	14-AUG-2000	200005-0225755
PR	18-AUG-2000	200005-0226271
PR	18-AUG-2000	200005-0226268
PR	22-AUG-2000	200005-0226866
PR	22-AUG-2000	200005-0227182
PR	23-AUG-2000	200005-0227009
PR	30-AUG-2000	200005-0228924
PR	01-SEP-2000	200005-0229287
PR	01-SEP-2000	200005-0229343
PR	01-SEP-2000	200005-0229344
PR	05-SEP-2000	200005-0229345
PR	05-SEP-2000	200005-0229509
PR	05-SEP-2000	200005-0229513
PR	06-SEP-2000	200005-0230437
PR	06-SEP-2000	200005-0230438
PR	08-SEP-2000	200005-0231242
PR	08-SEP-2000	200005-0231243
PR	08-SEP-2000	200005-0231244
PR	08-SEP-2000	200005-0231244
PR	08-SEP-2000	200005-0231413
PR	08-SEP-2000	200005-0231414
PR	08-SEP-2000	200005-0232080
PR	08-SEP-2000	200005-0232081
PR	12-SEP-2000	200005-0233196
PR	14-SEP-2000	200005-0233339
PR	14-SEP-2000	200005-0233398
PR	14-SEP-2000	200005-0233399
PR	14-SEP-2000	200005-0233400
PR	14-SEP-2000	200005-0233401
PR	14-SEP-2000	200005-0233603
PR	14-SEP-2000	200005-0233064
PR	14-SEP-2000	200005-0233065
PR	21-SEP-2000	200005-0234223
PR	21-SEP-2000	200005-0234274
PR	21-SEP-2000	200005-0234597
PR	25-SEP-2000	200005-0234998
PR	25-SEP-2000	200005-0235484
PR	27-SEP-2000	200005-0235834
PR	27-SEP-2000	200005-0235836
PR	29-SEP-2000	200005-0235837
PR	29-SEP-2000	200005-0236367
PR	29-SEP-2000	200005-0236368
PR	29-SEP-2000	200005-0236369
PR	29-SEP-2000	200005-0236370
PR	02-OCT-2000	200005-0236802
PR	02-OCT-2000	200005-0237037
PR	02-OCT-2000	200005-0237038
PR	02-OCT-2000	200005-0237039
PR	02-OCT-2000	200005-0237040
PR	13-OCT-2000	200005-0239355
PR	13-OCT-2000	200005-0239357
PR	20-OCT-2000	200005-0240960
PR	20-OCT-2000	200005-0241221
PR	20-OCT-2000	200005-0241221
PR	20-OCT-2000	200005-0241785
PR	20-OCT-2000	200005-0241786
PR	20-OCT-2000	200005-0241787
PR	20-OCT-2000	200005-0241808
PR	08-NOV-2000	200005-0246475
PR	08-NOV-2000	200005-0246476
PR	08-NOV-2000	200005-0246477
PR	08-NOV-2000	200005-0246478
PR	08-NOV-2000	200005-0246523
PR	08-NOV-2000	200005-0246524

[illegible]

and determined the sequences of interest, particular regions can be

KM Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KM
KM
KM ligase; hyperproliferative disorder; immunodeficiency disorder;
KM autoimmune disorder; neurological disorder; metabolic disorder;
KM inflammatory disorder; cardiovascular disorder; reproductive disorder;
KM blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KM nephrotropic; anticoagulant.
KM

PR 14-AUG-2000; 20000US-0225447.
PR 14-AUG-2000; 20000US-0225757

PR 14-AUG-2000; 2000US-0225759.


```
QY      1 HPELVKVAEDQLGARVGIETDLSGELLSPREPFPNMSTFKVLLCGAVLSRID 60
Db      28 HPELVKVAEDQLGARVGIETDLSGELLSPREPFPNMSTFKVLLCGAVLSRID 87
QY      61 AGQEQLRRIHYSNDLVESPYTEKHLTGKMTYRE 96
Db      88 AGQEQLRRIHYSNDLVESPYTEKHLTGKMTYXE 123

RESULT 6
AAM90871
ID AAM90871 standard; Protein; 129 AA.
AC AAM90871;
DX 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen SEQ ID NO:18464.
KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
PN M0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
```

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK63652.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Claim 11; SEQ ID NO 18464; 3071pp + Sequence listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 129 AA;
Query Match 35.4%; Score 477; DB 22; Length 129;
Best Local Similarity 96.9%; Pred. No. 8.8e-42;
Matches 93; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HPELVKVKDAEDQAGARVGYIELDLNSGELIESFRPERPPMSTFVLLCGAVLSRID 60
DB 28 HPELVKVKDAEDQAGARVGYIELDLNSGKILIESRPERPPMSTFVLLCGAVLSRID 87
OY 61 AGQDGLRRIRHSQNDLVEYSPVTEKHILTDGKTYRE 96
DB 88 AGQDGLRRIRHSQNDLVEYSPVTEKHILTDGKTYRE 123

RESULT 7
ABG27917
ID ABG27917 standard; Protein; 182 AA.
XX
AC ABG27917;

XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27908.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS92104.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 20; SEQ ID NO 58276; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
SQ Sequence 182 AA;
Query Match 35.4%; Score 477; DB 22; Length 182;
Best Local Similarity 87.3%; Pred. No. 1.5e-41;
Matches 96; Conservative 4; Mismatches 6; Indels 4; Gaps 1;

OY 1 HPELVKVKDAEDQAGARVGYIELDLNSGELIESFRPERPPMSTFVLLCGAVLSRID 60
DB 24 HPELVKVKDAEDQAGARVGYIELDLNSGKILIESRPERPPMSTFVLLCGAVLSRID 83
OY 61 AGQDGLRRIRHSQNDLVEYSPVTEKHILTDGKTYRELCASAITYSDNTAA 110
DB 84 AGQDGLRRIRHSQNDLVEYSPVTEKHILTDGKTYRE---LRRSPSDNTAS 129

RESULT 8
AAB59052
ID AAB59052 standard; Protein; 94 AA.

```

XX AC AAB59052;
XX XX
XX 27-MAR-2001 (first entry)
XX DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX XX
XX KW Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX KW neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX KW antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
XX KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX KW Addison's disease; allergic; autoimmune haemolytic anaemia;
XX KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX KW cardiovascular disorder; wound healing; neurological disease.
XX OS Homo sapiens.
XX PN WC020055173-A1.
XX XX
XX PD 21-SEP-2000.
XX XX
XX PF 08-MAR-2000; 2000WO-US05881.
XX XX
XX PR 12-MAR-1999; 99US-0124270.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Rosen CA, Ruben SM;
XX XX
XX DR WPI; 2000-611515/58.
XX DR N-PSDB; AAF21955.
XX XX
XX PT New human breast and ovarian cancer associated gene sequences and the
XX PT polypeptides encoded by these genes, useful in the prevention,
XX PT treatment and diagnosis of cancer, immune disorders, cardiovascular
XX PT disorders and neurological diseases -
XX XX
XX PS Claim 11; Page 1228; 1299pp; English.
XX XX
XX CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX CC associated with breast and ovarian cancer. Included in the invention are
XX CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX CC isolation and characterisation of the DNA and protein sequences of the
XX CC invention. The breast and ovarian cancer associated DNA, protein, agonist
XX CC or antagonist sequences exhibit cytostatic; immunosuppressive;
XX CC neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX CC antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
XX CC antibacterial; antifungal; antiparasitic and cardiant activity. The
XX CC polynucleotide and protein sequences are used in the diagnosis of cancer,
XX CC particularly breast and ovarian cancer. The nucleic acid sequences,
XX CC proteins, agonists and antagonists may also be used in the diagnosis,
XX CC prevention and treatment of immune disorders e.g. Addison's disease,
XX CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX CC arthritis and ulcerative colitis; cardiovascular disorders such as
XX CC myocardial ischaemia; wound healing; neurological diseases such as
XX CC cerebral anoxia and epilepsy; and infectious diseases.
XX XX
SQ Sequence 94 AA:
Query Match 33.9%; Score 457; DB 21; Length 94;
Best Local Similarity 97.8%; Pred. No. 6.9e-40;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

XX AC AAB60628
XX XX
XX ID AAB60628 standard; Protein; 159 AA.
XX XX
XX AC AAB60628;
XX XX
XX DT 25-MAR-2003 (updated)
XX DT 17-JUN-1991 (first entry)
XX XX
XX DE Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX XX
XX KW Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX KW fusion protein; beta-lactamase.
XX XX
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT Protein 1..96
XX FT /label= beta-lactamase
XX FT Misc-difference 97..100
XX FT /label= adaptor
XX FT Protein 101..159
XX FT /label= beta-urogastrone
XX XX
XX DE D5323634-A.
XX XX
XX PN 09-JAN-1986.
XX XX
XX PD 02-JUL-1985; 85DE-3523634.
XX XX
XX PR 02-JUL-1984; 84JP-0137691.
XX XX
XX PA (EART) EARTH CHEM CO LTD.
XX XX
XX PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
XX PI Matsushiro S;
XX XX
XX DR WPI; 1986-015031/03.
XX DR N-PSDB; AAN60632.
XX XX
XX PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX PT and transformed cells contg. it.
XX XX
XX PS Disclosure; Page 59-61; 92pp; German.
XX XX
XX CC The fusion protein is less easily degraded by proteases and so
XX CC protects beta-urogastrone and beta-lactamase collects in the periplasm
XX CC of E.coli. It is therefore easy to collect and purify the product.
XX CC Beta-urogastrone is the hormone of the salivary glands which suppresses
XX CC stomach acid secretion and promotes cell growth, so is useful for
XX CC treating ulcers and wounds. Previously the product was obt'd. only
XX CC in small amts. from human urine.
XX CC See also AAN60628, and 30-32.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX XX
SQ Sequence 159 AA:
Query Match 30.0%; Score 404; DB 7; Length 159;
Best Local Similarity 68.3%; Pred. No. 5.3e-34;
Matches 86; Conservative 9; Mismatches 19; Indels 12; Gaps 2;

```

Db 140 WVG YIG 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein; 101 AA.

AC ABG27935;

DT 18-FEB-2002 (flrst entry)

Novel human diagnostic protein #27926.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS92122.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutation/
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 58294; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. Abc00010-AB30317 represent novel human CC diagnostic amino acid sequences of the invention. CC CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at [ftp.wipo.int/pub/published.pct_sequences](http://wipo.int/pub/published.pct_sequences).

... Sequence 101 AA;

Query Match	27.2%	Score 366:	DB 22:	Length 101:
Best Local Similarity	98.6%	Pred. No. 2.6e-30:		
Matches 73: Conservative	0:	Mismatches 1:	Indels 0:	Gaps 0:

DY 154 DTTTPVAMATTLRKLTTCGELLTLASROQLIDMEADKVAGPLIRSLAPAGWFIADKSGAG 213
|||||
Db 10 DTTMPVAMATTLRKLITCELLTLASROQLIDMEADKVAGPLIRSLAPAGWFIADKSGAG 69

```

OY      214 ERGSRGIIALGPD 227
        |||||
Db      70  ERGSRGIIALGPD 83

```

RESULT 11	
ABG27919	
ID	ABG27919 standard; Protein; 88 AA.

AC ABG27919;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #27910.

Human; chromosome mapping; gene mapping; gene therapy; forensic food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dymanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-P\$DB; AAS92106.

PT New isolated polynucleotide and encoded polypeptides, useful in PT diagnostics, forensics, gene mapping, identification of mutations PT responsible for genetic disorders or other traits and to assess PT biodiversity -

PS Claim 20; SEQ ID No 58278; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Seq	Sequence	88 AA
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Query Match	24.7%	Score 333	DB 22	Length 88
Best Local Similarity	81.2%	Pred. No. 6.1e-27		
Matches 69: Conservative	3	Mismatches	3	Indels 10; Gaps 1

4 TLVKKVDAEDOLCARVGYIELDLSNGEILSFREPERFPMAMSTFKLLCGAVLSRIDAGQ 63

DB 1 TLVVKDAEDQIGARVGYIELDLSNGKILIESFRPERPRMSTFVLLCGAVLSRIDAGQ 60
 QY 64 EQLGRRI-----HYSONDLV 78
 XX ||||| | | | | |
 DB 61 EQLGRRIITLNMWTLSHOSKSL 85

RESULT 12

AAP60626
 ID AAP60626 standard; Protein; 119 AA.

AC AAP60626;

DT 25-MAR-2003 (updated)
 DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

KM Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 KM fusion protein; beta-lactamase.

OS Synthetic.

FH Key
 FT Protein 1.63 Location/Qualifiers
 FT /Label- beta-lactamase
 FT Misc-difference 64..66
 FT /Label- adaptor
 FT Protein 67..119
 FT /Label- beta-urogastrone

DE3523634-A.

PD 09-JAN-1986.

PF 02-JUL-1985; 85DE-3523634.

PR 02-JUL-1984; 84JP-0137691.

PA (EART) EARTH CHEM CO LTD.

PI Aoki S, Ohgaki H, Horinaka A, Hiramatsu H, Komoto S, Nishimura A;
 PI Matsushiro S;

DR WPI; 1986-015031/03.

DR N-PSDB; AAN60630.

PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 PT and transformed cells contg. it.

PS Disclosure; Page 55; 92pp; German.

CC The fusion protein is less easily degraded by proteases and so
 CC protects beta-urogastrone and beta-lactamase collects in the periplasm
 CC of E.coli. It is therefore easy to collect and purify the product.
 CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 CC stomach acid secretion and promotes cell growth, so is useful for
 CC treating ulcers and wounds. Previously the product was obt. only
 CC in small ants. from human urine.
 CC See also AAN60628, and 30-32.
 CC (Updated on 25-MAR-2003 to correct PA field.)

SQ Sequence 119 AA;

Query Match 14.8%; Score 199; DB 7; Length 119;

Best Local Similarity 97.5%; Pred. No. 1e-12; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPETLVKVKDAEDQIGARVGYIELDLSNGKILIESFRPER 40
 DB 24 HPETLVKVKDAEDQIGARVGYIELDLSNGKILIESFRPER 63

RESULT 13
 AA92783
 ID AA92783 standard; Protein; 51 AA.

AC AA92783;

DT 29-AUG-2000 (first entry)

DE Vtgs-Beta-lactamase fusion protein (partial).

KM Vtgs; vitellogenin; secretory signal sequence; gene expression;
 KM oestrogen receptor binding protein; systemic circulation;
 KM beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

FH Key
 FT Peptide 1..21 Location/Qualifiers
 FT /Label- Vtgs
 FT Cleavage-site 15..16
 FT Protein 22..51
 FT /note= "beta-lactamase mature protein"

WO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

PA (UYSI-) UNIV SINGAPORE NAT.
 PA (LAWT/) LAM T J.

PI Ding JL, Tan NS, Ho B;

DR WPI; 2000-365615/31.

DR N-PSDB; AAA28507.

PT Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

PS Example 6; Fig 14A; 73pp; English.

CC A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pBADVtgsbLactKana was
 CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an estrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

SQ Sequence 51 AA;

Query Match 10.5%; Score 141; DB 21; Length 51;

Best Local Similarity 96.8%; Pred. No. 3.6e-07; Mismatches 0; Indels 0; Gaps 0;

QY 2 PETLVKVKDAEDQIGARVGYIELDLSNGK 30
 DB 23 PETLVKVKDAEDQIGARVGYIELDLSNGK 51

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds

(without alignments)
938,485 Million cell updates/sec

Title: SEQ2_30E_157T

Sequence: 1 HPETIVKVKAEQDLARVC.....TMDERNQIAIGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCUTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.5	5.6	170	US-08-858-207A-519	Sequence 519, App
2	74.5	5.5	124	US-08-811-731A-202	Sequence 202, App
3	72.5	5.4	197	US-09-252-991A-24241	Sequence 24241, A
4	71	5.3	159	US-08-991-890-4	Sequence 4, Appl
5	69	5.1	169	US-09-328-352-4491	Sequence 4491, Ap
6	65.5	4.9	158	US-09-010-809-19	Sequence 19, Appl
7	65	4.8	150	US-09-239-909-2	Sequence 2, Appl
8	65	4.8	181	US-08-482-142-195	Sequence 195, App
9	65	4.8	181	US-08-478-572-195	Sequence 195, App
10	64.5	4.8	181	US-08-484-296-195	Sequence 195, App
11	64.5	4.8	178	US-09-252-991A-29942	Sequence 29942, A
12	64	4.7	203	US-08-624-677A-2	Sequence 2, Appl
13	63.5	4.7	144	US-08-225-480-4	Sequence 4, Appl
14	63.5	4.7	144	US-09-118-445-4	Sequence 4, Appl
15	63	4.7	174	US-08-557-122A-6	Sequence 6, Appl
16	63	4.7	174	US-09-262-666-6	Sequence 6, Appl
17	63	4.7	200	US-08-557-122A-12	Sequence 12, Appl
18	63	4.7	200	US-09-262-666-12	Sequence 12, Appl
19	62	4.6	132	US-09-252-991A-22681	Sequence 22681, A
20	61.5	4.6	142	US-09-345-473E-31	Sequence 31, Appl
21	61	4.5	189	US-08-671-548C-48	Sequence 48, Appl
22	61	4.5	204	US-09-252-991A-17611	Sequence 17611, A
23	60.5	4.5	103	US-09-732-210-1282	Sequence 1282, Ap
24	60.5	4.5	167	US-08-690-849-2	Sequence 2, Appl
25	60.5	4.5	167	US-09-004-053-2	Sequence 2, Appl
26	60	4.5	158	US-09-107-532A-4218	Sequence 4218, Ap
27	59.5	4.4	146	US-09-134-001C-5269	Sequence 5269, Ap

28	59.5	4.4	178	US-09-220-731-24	Sequence 24, Appl
29	59.5	4.4	178	US-09-242-999-24	Sequence 24, Appl
30	59.5	4.4	144	US-09-252-991A-25578	Sequence 25578, A
31	59	4.4	171	US-09-107-532A-3979	Sequence 3979, Ap
32	59	4.4	180	US-09-194-905-5	Sequence 5, Appl
33	59	4.4	199	US-09-252-991A-30363	Sequence 30363, A
34	58.5	4.3	167	US-09-328-352-6278	Sequence 6278, Ap
35	58.5	4.3	168	US-09-413-814-87	Sequence 87, Appl
36	58.5	4.3	203	US-09-252-991A-24921	Sequence 24921, A
37	58	4.3	168	US-08-451-947-6	Sequence 30622, A
38	58	4.3	168	US-08-424-826A-6	Sequence 6, Appl
39	58	4.3	168	US-08-928-694-6	Sequence 6, Appl
40	58	4.3	168	US-08-450-842-6	Sequence 6, Appl
41	58	4.3	168	US-08-451-390-6	Sequence 6, Appl
42	58	4.3	168	US-08-451-390-6	Sequence 6, Appl
43	58	4.3	168	US-09-252-991A-31347	Sequence 31347, A
44	58	4.3	183	US-09-252-991A-29325	Sequence 29325, A
45	58	4.3	187		

ALIGNMENTS

RESULT 1
US-08-858-207A-519
Sequence 519, Application US/08858207A

Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Nicholas, Richard

TITLE OF INVENTION: No 6348328e1 Compounds

NUMBER OF SEQUENCES: 552

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858, 207A

FILING DATE: 09-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017670

FILING DATE: 14-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50475

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 519:

SEQUENCE CHARACTERISTICS:

LENGTH: 170 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6348328e

US-08-858-207A-519

Query Match

Best Local Similarity

5.68: Score 75.5; DB 4; Length 170;

22.18; Pred. No. 2.2;


```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-991-890-4

```

```

Query Match 5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.3;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

```

```

DB 49 VILCGAVLRIDAGD-----QLGRRIHSQN-----DLVEYSPYTERH 87
DB 12 LLLCGAVF--VSPDSQIHAEFGRRHHHHHGGSGAELRGCGPRFGKHLSTCPMPREKT 69
DB 88 LVDGMTVRELCSAATMSDNTAANLLTTIGPKELTAFLNMGDHYTLDRPREPLNLA 147
DB 70 FT-----TTPGCVLLES-GRPKENVSTNNKCGAL-----GTTSEF 105
DB 148 IPNDERDTPVAMA-TTLRKL 169
DB 106 IPNLSPELKKPLSEGGPSLKKII 128

```

```

RESULT 5
US-09-328-352-4491
Sequence 4491, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4491
LENGTH: 169
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

```

```

Query Match 5.1%; Score 69; DB 4; Length 169;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

```

```

DB 118 GSPKELTAFLNMGDHYTLDRWEPELNAIPNDERDT-----TPVAATTLRKL 169
DB 5 GGEINTMKMLAQON---RFDQWVEDIYVL---DRLLTVREHIAVFTTFVYVYVIGYS 58
DB 170 TGBELLTLASRQ-----LIDWMEADVAGPLRLSALPAGWFLADKSGAGER--GSRG 219
DB 59 LMKHSLAEQQQKRLNDLKDLMVYMQSNV-----TKMPANELDEFGSKIQRYVAQQG 112
DB 220 ITAALGPDGKPSRIYVYTTGQATDENRQIAETGASLIR 261
DB 113 LVVSSQONGEQLQIV--THONATILANFLTLQADGKLSIQK 152

```

```

RESULT 6
US-09-010-809-19
Sequence 19, Application US/09010809B
Patent No. 6090601
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
APPLICANT: Bellach, Mary C.
TITLE OF INVENTION: Epothilone Polyketide Synthases and Encoding DNA
FILE REFERENCE: 30062-20020.00
CURRENT APPLICATION NUMBER: US/09/010,809B
CURRENT FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 158
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-010-809-19

```

```

Query Match 4.9%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 26;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

```

```

DB 106 DNTANLLLT-----TIGPKELTAFLNMGDHYT---RIDRREP-----ELNFAIPDERD 154
DB 21 NHDAKHLILTSRQASAPGADYLRSELALGASVTLAADVADPRALKDLNINP----- 75
DB 155 TTPVAMATTIRKLTLGELLTLASRQQLDWMHEADVAGPLRSA-----LPAGNF 205
DB 76 SAHPVAAYVHAASVLDGLGANSLERI-----DRVFAPKIDAAWHLQLTQDCKPLAIF 129
DB 206 IADKSGAGERSRG 219
DB 130 ILFSSVAGVLTSSSG 143

```

```

RESULT 7
US-09-239-909-2
Sequence 2, Application US/09239909
Patent No. 6284952
GENERAL INFORMATION:
APPLICANT: Kumho Petrochemical Co. Ltd.
TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
FILE REFERENCE: P99-2-6
CURRENT APPLICATION NUMBER: US/09/239,909
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: EP 99300136.1
EARLIER FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: KOPATIN 1.0
SEQ ID NO 2
LENGTH: 150
TYPE: PRT
ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

```

```

Query Match 4.8%; Score 65; DB 3; Length 150;
Best Local Similarity 27.9%; Pred. No. 27;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

```

```

DB 73 SONDVEYSPYTERKHLTDS---MTVRELCSAATMSDNTANLLTTI-----GGPKR 122
DB 6 SEQIYDFKEAFGLFDKDCGCTVEELATVIRSLDQPTREELQDMISEVDADGNGTIE 65
DB 123 LNAFLNMGDHYTLDRWEPELNAIPNDERDTPVAMATTLRKL--TGBELLTLASRQ 180
DB 66 FDEFSLAKKAYKQDA-BEELNEAFKVDKQNGTIS-ASELRHVMINLGEKLTDEEVE 123
DB 181 QLIDWMEAD 189
DB 124 QMI--KEAD 130

```

```

APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-Chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,572
FILING DATE: 07-June-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,605
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-572-195

Query Match 4.8%; Score 65; DB 2; Length 181;
Best Local Similarity 20.5%; Pred. No. 36;
Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5;

QY 4 TLVAVKADQDLGARVGT--TELDLNGSEILIESRPERFRPMSTFKVLGCAVLSRIDA 61
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 17 THAKIRVVAATESAYLAVRNTSLDLSQELVDEYIOHNGVQESYRYV----- 65
Y 62 GQGLDGRRIHSQ-----NDLVESYPTKHLTLDGTVRELCSAATMSDNTANILL 114
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 66 AREQSCRPNQOQLEAVFEANQNTKATIEIKASIDGLEV----- 105
Y 115 TTIGGPELTAFLNMGDHYTRLDRWEPDLNLAIPND---ERDTTPVAM 161
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
DB 106 --IIGINDLDAFRHYDGRITIIQRONGVOSINGNAPAEIDLQNTVPIRM 154

RESULT 10
US-08-484-296-195
Sequence 195. Application US/08484296
Patent No. 6268491
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-Chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian

```

QY 54 AVLSRIDAGOEQLGRRIHYSQNDLVEYSPVTEKHLTDGMTVREICSA - ITNSDNTAAN 111

Db 33 AGVSNIDGDDDAAGNPDP-----SDVTDAITDGEPRVYSGKPHITQKSLIK 83
 QY 112 LLTTIGCPKELTAFLHNGDHV-TRLDRNEPELNEAIPNDERDT-----T 157
 Db 84 LAMPVVA---LTSYL--VADRVLPELTSNEEGTESIPKKRYKAVGIALVAAAAFA 138
 QY 158 PVMATLRLKLLGELLTLASROOLDIMWEADYVAGPLRSALP 201
 Db 139 GLGLATFRHFVPRKSKTVASEDSALGNSEQYVEGTVNGSSDP 182

RESULT 13
 US-08-225-480-4
 ; Sequence 4, Application US/08225480
 ; Patent No. 5817793
 ; GENERAL INFORMATION:
 ; APPLICANT: LEVY, STUART B.
 ; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
 ; TITLE OF INVENTION: ASSAYS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/225.480
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; PRIOR APPLICATION NUMBER: US 07/938,085
 ; FILING DATE: 28-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: T0359/7003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-720-3500
 ; TELEFAX: 617-720-2441
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 144 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-225-480-4

Query Match 4.7%; Score 63.5; DB 2; Length 144;
 Best Local Similarity 29.5%; Pred. No. 37;
 Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;
 QY 66 LGRRIHY--SOND--LVEY--SP--VTEKHILTDGMYRELCSAATMSDNTAANLLTTIG 118
 Db 14 LGRILHVNOKRDLNELYSLPDITAAQFVLCISIR--CAACIT----- 56
 QY 119 GKELTAFLH-NMGDHYTRLDR-----WEPELNEAIPNDERD-----TTPVAMATTLRK 167
 Db 57 -PVELKATVLDGALTRMLDRLVCKGWVERLPN--PNDRKGYLVKLTGGAALICGCHQ 113
 QY 168 LITGELLTLASROOLDIMWEADKYA--GPLRSALP 201
 Db 114 LVGQDL-----HQLTKMLTADEVATLEYLLKVL 144

US-09-118-445-4
 ; Sequence 4, Application US/09118445
 ; Patent No. 6391545
 ; GENERAL INFORMATION:
 ; APPLICANT: LEVY, STUART B.
 ; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
 ; TITLE OF INVENTION: ASSAYS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
 ; STREET: 600 ATLANTIC AVENUE
 ; CITY: BOSTON
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/118.445
 ; FILING DATE: 17-Jul-1998
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/225,480
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: GATES, EDWARD R.
 ; REGISTRATION NUMBER: 31,616
 ; REFERENCE/DOCKET NUMBER: T0359/7003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-720-3500
 ; TELEFAX: 617-720-2441
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 144 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 ; US-09-118-445-4

Query Match 4.7%; Score 63.5; DB 4; Length 144;
 Best Local Similarity 29.5%; Pred. No. 37;
 Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;
 QY 66 LGRRIHY--SOND--LVEY--SP--VTEKHILTDGMYRELCSAATMSDNTAANLLTTIG 118
 Db 14 LGRILHVNOKRDLNELYSLPDITAAQFVLCISIR--CAACIT----- 56
 QY 119 GKELTAFLH-NMGDHYTRLDR-----WEPELNEAIPNDERD-----TTPVAMATTLRK 167
 Db 57 -PVELKATVLDGALTRMLDRLVCKGWVERLPN--PNDRKGYLVKLTGGAALICGCHQ 113
 QY 168 LITGELLTLASROOLDIMWEADKYA--GPLRSALP 201
 Db 114 LVGQDL-----HQLTKMLTADEVATLEYLLKVL 144

RESULT 15
 US-08-557-122A-6
 ; Sequence 6, Application US/08557122A
 ; Patent No. 5879664
 ; GENERAL INFORMATION:
 ; APPLICANT: Hjort, Carsten Mølland
 ; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
 ; NUMBER OF SEQUENCES: 38
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NO. 58796640 No. 58796640sk of No. 5879664th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York

```

; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-557-122A-6

Query Match      4.7%; Score 63; DB 2; Length 174;
Best Local Similarity 23.9%; Pred. No. 56;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KVDADDEQLGAR-VGYIELDLNSGEILSEFRPEERFPMSTFVILGAVLSRIDAGOEQ 65
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 67 KYEQATELKEKNIPLYKVCETEEFALCRDQGVGGYPTLKIFRGL-----DAVKPY 117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 66 LGRR-----IHYS-QNDLVEYSPYTEKHLTDGWTVRRLCSAALTMSDNTANLLFTT 116
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 118 QGARQTEAIVSYWKOSLPAAVSPTEPNLEIKTKMDKIIVIGYIASDDQTANDIFTT 174
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: September 10, 2003, 12:31:05
Job time : 12.8571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E_157T

Perfect score: 1348

Sequence: 1 HPEFLVKVKAEDQIGARVG.....TMDERNQIAEIGASLIRHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues 278642

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
- 16: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	457	33.9	94	US-10-102-806-760	Sequence 760, App
2	75	5.6	202	US-10-127-816-9	Sequence 9, App1
3	75	5.6	202	US-10-142-717-12	Sequence 12, App1
4	70.5	5.2	193	US-09-805-354-8	Sequence 8, App11
5	70.5	5.2	193	US-10-144-259-8	Sequence 8, App11
6	70	5.2	206	US-10-156-761-13867	Sequence 13867, A
7	69.5	5.2	184	US-10-101-464A-765	Sequence 765, App
8	69	5.1	149	US-10-156-761-8136	Sequence 8136, App
9	67.5	5.0	149	US-10-233-926-4	Sequence 4, App11
10	67.5	5.0	187	US-09-948-018-8	Sequence 8, App11
11	67	5.0	178	US-09-738-626-5680	Sequence 5680, App
12	67	5.0	194	US-10-156-761-8623	Sequence 8623, App
13	66	4.9	202	US-10-127-816-11	Sequence 11, App1
14	65.5	4.9	191	US-09-764-870-390	Sequence 390, App
15	65.5	4.9	191	US-10-125-540-390	Sequence 390, App

16	64.5	4.8	189	9	US-09-815-242-5862	Sequence 5862, App
17	64.5	4.8	189	9	US-09-815-242-12979	Sequence 12979, A
18	64.5	4.8	191	15	US-10-156-761-8978	Sequence 8978, App
19	64	4.7	162	10	US-09-738-626-4796	Sequence 4796, App
20	64	4.7	179	10	US-09-764-868-757	Sequence 757, App
21	64	4.7	179	15	US-10-106-898-4858	Sequence 4858, App
22	63.5	4.7	202	15	US-10-189-346-12	Sequence 12, App1
23	63.5	4.7	88	15	US-10-100-252-6	Sequence 6, App11
24	63.5	4.7	144	15	US-10-131-406-4	Sequence 4, App11
25	63.5	4.7	206	10	US-09-738-626-5425	Sequence 5425, App
26	62.5	4.6	190	15	US-10-156-761-9507	Sequence 9507, App
27	62.5	4.6	195	15	US-10-156-761-12656	Sequence 12656, App
28	62	4.6	174	9	US-09-864-761-35777	Sequence 35777, A
29	62	4.6	191	11	US-09-927-827-64	Sequence 64, App1
30	62	4.6	202	15	US-10-156-761-13460	Sequence 13460, A
31	61.5	4.5	142	10	US-09-862-027-31	Sequence 31, App1
32	61.5	4.5	196	10	US-09-738-626-6624	Sequence 6624, App
33	61.5	4.6	200	15	US-10-156-761-12006	Sequence 12006, A
34	61	4.5	189	9	US-09-764-870-523	Sequence 523, App
35	61	4.5	189	15	US-10-125-540-523	Sequence 523, App
36	61	4.5	197	10	US-09-803-661-5	Sequence 5, App11
37	61	4.5	197	15	US-10-300-827-5	Sequence 5, App11
38	60.5	4.5	160	12	US-09-882-227-230	Sequence 230, App
39	60.5	4.5	202	15	US-10-189-346-16	Sequence 16, App1
40	60	4.5	185	15	US-10-101-464A-616	Sequence 616, App
41	60	4.5	192	10	US-09-738-626-4637	Sequence 4637, App
42	59.5	4.4	182	15	US-10-156-761-12056	Sequence 12056, A
43	59.5	4.4	196	15	US-10-219-220-265	Sequence 265, App
44	59.5	4.4	206	9	US-09-741-669-461	Sequence 461, App
45	59.5	4.4	206	9	US-09-815-242-10335	Sequence 10335, A

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102, 806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925, 298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124, 270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-760

Query Match 33.9%; Score 457; DB 15; Length 94;
Best Local Similarity 97.8%; Pred. No. 1.3e-39;
Matches 86; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 75 NDIVEYSPVTEKHLTDGMTVRELCSAATNSDNTANLITTTGGPRELTAFLHNGDHV 134
DB 1 NDIVEYSPVTEKHLTDGMTVRELCSAATNSDNTANLITTTGGPRELTAFLHNGDHV 60

```
QY      135 TRLDWEPELNPAIPNDERDITTPVAMATT 164  
        |||||  
Db      61  TRLDWEPELNPAIPNDERXTTMPVAMATT 90
```

RESULT 2
US-10-127-816-9

```

1  Sequence 9, Application US/20127816
2  Publication No. US20030104416A1
3  GENERAL INFORMATION:
4  APPLICANT: Sheppard, Paul O.
5  APPLICANT: Fox, Brian A.
6  APPLICANT: Kluehner, Kevin M.
7  APPLICANT: Taft, David W.
8  APPLICANT: Kindsvogel, Wayne R.
9  TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
10 FILE REFERENCE: 01-17
11 CURRENT APPLICATION NUMBER: US/10/127,816
12 CURRENT FILING DATE: 2002-04-19
13 PRIOR APPLICATION NUMBER: US 60/285,408
14 PRIOR FILING DATE: 2001-04-20
15 PRIOR APPLICATION NUMBER: US 60/286,482
16 PRIOR FILING DATE: 2001-04-25
17 PRIOR APPLICATION NUMBER: US 60/341,050
18 PRIOR FILING DATE: 2001-10-22
19 PRIOR APPLICATION NUMBER: US 60/341,105
20 PRIOR FILING DATE: 2001-10-22
21 PRIOR APPLICATION NUMBER: US 09/895,834
22 PRIOR FILING DATE: 2001-06-29
23 PRIOR APPLICATION NUMBER: US 60/285,424
24 PRIOR FILING DATE: 2001-04-20
25 NUMBER OF SEQ ID NOS: 59
26 SOFTWARE: FastSeq for Windows Version 4.0
27 SEQ ID NO: 9
28 LENGTH: 102
29 TYPE: prt
30 ORGANISM: Homo sapiens
31 US-10-127-816-9

```

Query Match 5.68; Score 75; DB 15; Length 203;
Best Local Similarity 21.48; Pred. No. 13;
Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

[illegible]

US-10-142-717-12
 RESULT 3
 Sequence 12, Application US/10142717
 PUBLICATION NO. US20030104579A1
 GENERAL INFORMATION:
 APPLICANT: Immunex Corporation
 APPLICANT: Baum, Peter R
 APPLICANT: Mosley, Bruce A
 APPLICANT: Ketchum, Randal R
 APPLICANT: Taylor, Scott L
 TITLE OR INVENTION: CYTOKINE POLYPEPTIDES
 FILE REFERENCE: 3282-A

```

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 12

```

Query Match	5.68	Score	75	DB	15	Length	202
Best Local Similarity	21.48	Pred. No.	13				
Matches	40	Conservative	26	Mismatches	67	Indels	54
						Gaps	7

[illegible]

RESULT 4
US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION

Query Match	5.28;	Score 70.5;	DB 11;	Length 193;
Best Local Similarity	24.28;	Pred. No. 35;		
Matches	31;	Conservative	15;	Mismatches 35; Indels 47; Gaps 5

QY 59 DADGEGDGRGRHYSQNDLVESR--VTEKRLTGCTVRREICSAATIMSDNTAALLTTT 116
| | | | : | | | : | : | : |
Dbd 36 IGGPDIQVG-----VVQGEDVVEHFELNDYRSVKDYVAASHIEQR----- 77
| | | | : | : | : | : | : |
QY 117 IGGPELTAFLPHNMCDHTRLDKRE-----PETNEALPNDERD 154
| | | | : | : | : | : | : |
Dbd 78 -GGETTRAF---GIEPARSEAFQGGKRGAKRMIVITDGESHSPDEKVIQQSERD 192
| | | | : | : | : | : | : |
QY 155 TTTTPYAMA 162
| | | | : | : | : | : | : |
Dbd 133 NVTRYAVA 140

RESULT 5

US-10-144-259-8
; Sequence 8, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaud, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8

Query Match 5.2%; Score 70.5; DB 15; Length 193;
Best Local Similarity 24.2%; Pred. No. 35;
Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;
QY 59 IDAGOEOLGRIRHSONDLVEYSP--VTEKHLTDGNTVRELCSAITMSDNTAALLTTT 116
DB 36 IGPOQIOVG-----VVOYGEDVYHEFLNDYRSVKDVEAASHIEGR----- 77
QY 117 ICGREKLEAFHNNGDVTRLDRE-----PELNEAIPNDERD 154
DB 78 -GGERTETAF-----GIFASSEAFQKGRKAKVMIVITDGSHPDLEKVIQOSERD 132
QY 155 TTTVYAMA 162
DB 133 NVTXYAVA 140

RESULT 6
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13867
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867

Query Match 5.2%; Score 70; DB 15; Length 206;
Best Local Similarity 25.2%; Pred. No. 44;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;
QY 70 IHSONDLVEYSPVTEKHLTDGNTV-----RELCSAITMSDNTAALLTTT---G 118
: | | | | | : | | : | | : | | | | |

DB 1 VHASOG---AOGPTESAGEKYTMADAGDSFREVA-----NRSSALLNTAVLLSG 50
QY 119 GPKF-----LTAFLHNNGDVTRLD-----RW-----PELNEAIPNDE 152
DB 51 GDRHAADLLONALIKADRWSRIDEPEAVYQVLYROQYSRRRLKRWRELSSVAP--PE 109
QY 153 RDTTPVAMNTTIRKLLTGELLLASRQOLI-----DMEADKYAGPLRSALPAGMF 205
DB 110 ASTEPDASAEELRVGALARLTARQTVLYRYFEDLPEAD--VARILGSGVGTYS 167
QY 206 IADKSGAGERGSKITIALGP---DGKPSR 232
DB 168 THRSRLARLTAPELAAAGPADADGEPSS 197

RESULT 7
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000,102062
; CURRENT APPLICATION NUMBER: US/10/101,464A
; PRIOR FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-765

Query Match 5.2%; Score 69.5; DB 15; Length 184;
Best Local Similarity 24.7%; Pred. No. 42;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;
QY 41 FPMSTFXYL-----LCGAVLSRIDAGOEOLGRIRHSONDLVEYSPV---TEKHLTDG 91
DB 23 FCMSTERTLVYPYMLNGSVASRL-----RDSINGKPALDMPTRKRISLGS 67
QY 92 -----MTVRELCSAITMSDNTAALL-----TTIGREKLEAFHNNGDVTRLDNR-- 140
DB 68 MARGILTYLHQCQDPITIRDVKAAWILLDEYFEAVYGGFGAKLIDHDSHTVAVAGTV 127
QY 141 ---PELNEAIPNDERDITTPVAMNTTIRKLLTGEL---LTLASROO--LIDNM 186
DB 128 GHIAPEYLSFGOSSEK--TDVFGGIIULELITGOKALDPRRAANQGVMLDWY 179

RESULT 8
US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA

```

      TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
      FILE REFERENCE: 249-262
      CURRENT APPLICATION NUMBER: US/10/156,761
      CURRENT FILING DATE: 2002-05-29
      PRIOR APPLICATION NUMBER: JP 2001-204089
      PRIOR FILING DATE: 2001-05-30
      PRIOR APPLICATION NUMBER: JP 2001-272697
      PRIOR FILING DATE: 2001-08-02
      NUMBER OF SEQ ID NOS: 15109
      SEQ ID NO 8136
      TYPE: PRT
      LENGTH: 149
      ORGANISM: Streptomyces avermectinis
      US-10-156-761-8136

Query Match          5.1%; Score 69; DB 15; Length 149;
Best Local Similarity 30.9%; Pred. No. 35;
Matches 23; Conservative 13; Mismatches 24; Indels 20; Gaps 3;

QY      168 LNTGELLTL-----ASROQLDMMEDKVGPLRLSALPAGFIADKSG----- 211
       |||||||:::||||| | | | | | :|||:
DB      50 LLDIELLTITKLRLVAVSDAKKEKGIDWHEHD---PALSHADGGRELAENRRLREEI 105

QY      212 AGERGRGITIALGPDKPSR 232
       |||:::| | | | | | | | |
DB      106 AGLRAQAALPSPAEPAEKPER 126

RESULT 9
US-10-233-926-4
Sequence 4, Application US/10233926
Publication No. US20030131382A1
GENERAL INFORMATION:
APPLICANT: Allen, Stephen M.
APPLICANT: Kinney, Anthony J.
APPLICANT: Falco, S. Carl
TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
FILE REFERENCE: BB1419 US NA
CURRENT APPLICATION NUMBER: US/10/233,926
CURRENT FILING DATE: 2002-09-03
PRIOR APPLICATION NUMBER: US/09/735,846
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Microsoft Office 97
SEQ ID NO 4
LENGTH: 149
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (18)
US-10-233-926-4

Query Match          5.0%; Score 67.5; DB 16; Length 149;
Best Local Similarity 21.4%; Pred. No. 50;
Matches 22; Conservative 23; Mismatches 33; Indels 25; Gaps 4;

QY      139 VAMATTKRL-----TGELITLASROQLDMMEDKVGGLLSALPAGFIADKSGA 212
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      14 LSLXSLPLMTLMADHAALAAPSSQEEDPEMKLEEGGDGV-----EVADRGSG 64

QY      213 GERGRGITIALGPDKPSRIYV-----IYTTSOATMDERNR 250
       | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      65 GCANAGCI-----PECRPTRIVADGIYDLFHRGHAKSLDAQAR 102

RESULT 10
US-09-948-018-8
Sequence 8, Application US/09948018
Patent No. US20020150977A1
GENERAL INFORMATION:
APPLICANT: Theill et al
TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
```

	FILE REFERENCE:	01017/3767			
	CURRENT APPLICATION NUMBER:	US/09/948_018			
	CURRENT FILING DATE:	2001-09-05			
	PRIOR APPLICATION NUMBER:	US 60/230_191			
	PRIOR FILING DATE:	2000-09-05			
	NUMBER OF SEQ ID NOS:	45			
	SOFTWARE:	PatentIn version 3.1			
	SEQ ID NO 8				
	LENGTH:	187			
	TYPED PRIMER:				
	ORGANISM:	Homo sapiens			
	US-09-948-018-8				
	Query Match	5.0%; Score 67.5;	DB 10; Length 187;		
	Best Local Similarity	22.4%; Pred. No. 69;			
	Matches	41; Conservative	22; Mismatches 57; Indels 63; Gaps 9;		
OY	112 LLTTTGGPKELTAFLHNMGDHYTRIDWEE-----LNEALPNDERTITPVAMAT-	163			
Dd	11 LLLLAAPPPPEAS-----OYCGLLEYWNPDNKCSCSLQRFGE-----PPCPALETG	58			
OY	164 -----TKRLKITGELLTLASR--COLIDMMEA-----DKVAGP-----L	195			
Dd	59 DTWKRASTLPPLSRSLSSLASPSLRLLDELEVELELVLDPERPGSGMAHTTHIA	118			
OY	196 LRSAIPACMFITADKSGAGEGRSRTIALMGDPGFSTR---IYYITYTGQAOTDENNRQIA	253			
Dd	119 AAYGPAAW-----STEPAYSILRPSPRLALIENVARREPSAISIGQLGTIIA	165			
OY	254 EIG 256				
Dd	166 QYG 168				
	RESULT 11				
	US-09-738-626-5680 Application US/09738626				
	Sequence 5680, Publication No. US20020197605A1				
	GENERAL INFORMATION:				
	APPPLICANT: NAKAGAWA, SATOSHI				
	APPPLICANT: MIZOGUCHI, HIROSHI				
	APPPLICANT: ANDO, SEIKO				
	APPPLICANT: HAYASHI, MIKIHO				
	APPPLICANT: OCHIAI, KEIRO				
	APPPLICANT: YOKOI, HARUHIKO				
	APPPLICANT: TATEISHI, NAOKO				
	APPPLICANT: SENOH, AKIHIO				
	APPPLICANT: IKEDA, MASATO				
	APPPLICANT: OKAI, AKIO				
	TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES				
	FILE REFERENCE: 249-125				
	CURRENT APPLICATION NUMBER: US/09/738,626				
	CURRENT FILING DATE: 2000-12-18				
	PRIOR APPLICATION NUMBER: JP 99/377484				
	PRIOR FILING DATE: 1999-12-16				
	PRIOR APPLICATION NUMBER: JP 00/159162				
	PRIOR FILING DATE: 2000-04-07				
	PRIOR APPLICATION NUMBER: JP 00/280988				
	PRIOR FILING DATE: 2000-08-03				
	NUMBER OF SEQ ID NOS: 7059				
	SOFTWARE: Patentln ver. 3.0				
	SEQ ID NO 5680				
	LENGTH: 178				
	TYPE: PRT				
	ORGANISM: Corynebacterium glutamicum				
	US-09-738-626-5680				
	Query Match	5.0%; Score 67;	DB 10; Length 178;		
	Best Local Similarly	24.8%; Pred. No. 73;			
	Matches	25; Conservative	17; Mismatches 37; Indels 22; Gaps 4;		
OY	171 GELLTASSROOLDIM-----MEADVAPGLRSALPACMTFADKSAGERSRGITIAA	223			
	:: :::: : :: :				

```

Db      61 GELPDMAEGBELNFGACITILEVSTPGVDNPL---TLPRH-----RRNRGLVA 107
QY      224 LGPDGKP--SRIVVYTTGSCATMDERNRQIAEIGASLIRK 262
      108 LDODGKRKRVARIGALNDAETHVLIERNKKLEVTTELIAH 148

RESULT 12
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 194;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

QY      202 AGWFINDKSGAGRGSGRG---TIALGPDG--KPSRIYVITTT 239
      75 SGRVTPRSGAGRGRTYEGAEYRTTAIGPLAVHEPVVAVVVT 118
Db      75 SGRVTPRSGAGRGRTYEGAEYRTTAIGPLAVHEPVVAVVVT 118

RESULT 13
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klueber, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match
Best Local Similarity 4.9%; Score 66; DB 15; Length 202;
Matches 41; Conservative 23; Mismatches 65; Indels 62; Gaps 8;

QY      50 LILGAVLSRIDAGQEQOLGRIRHYSQND--LVESPYTEKHL-----TDCNTREL----- 97
      17 LILAAVLTFOADPPVPVATRLPVEAKDCHIAQFKSLSPKEIQAFKAKAIEKRLLENDM 76
QY      98 -CSA-----AITMS--DNTAANLLTTIGGPKELATFLHN 129
      77 KCSHLISRAMDKLOLQOVRPKALQAEVALTKWENINDSALTITIGQPLHTLSHHS 136
Db      130 MGDHVT-----RLDRWEPLENEA---IPNDRPTTPVMAATLRALLGE 172
      137 OLQTCQLQATAPKPPPSRLSRWLRLQADSKETPGCEDSVT-----SNTFQLLRD 191
QY      173 LILTASRQDLT 183
      192 LKCVASGDQCV 202
Db

```

```

RESULT 14
US-09-764-870-390
; Sequence 390, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; PRIOR Application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-390

```

```

Query Match
Best Local Similarity 4.9%; Score 65.5; DB 9; Length 191;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;

QY      21 YIELDNSGELLESFREFEFPMMSTFKVLLCGAVLSRIDAGQEQOLGRIRHYSQNDLVEX 80
      48 YCRSLNDSEY-----DMFGDYDSEFNSFIAQVDDLQK-----YMO----- 85
QY      81 SPVTEKHLTDGMVRELCSAATMSDNTAANLLTTIGGKRELTAFLLH----- 128
      86 LPEHKHATDPAT-ENLCSESI-----KKKLSITITIGNTLEQTYKKHBNOSGYGCVTI 138
QY      129 -----NMGDVTTRLDRWEPE-----LNEAIPND 151
      139 EPGADLLYDVPSSQAIYFNKLQNSNDLGDHSHKDRDXXSSSHNTVNEELPRH 191
Db

```

```

RESULT 15
US-10-125-540-390

```

```

; Sequence 390, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT1A1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-125-540-390

Query Match      4.98; Score 65.5; DB 15; Length 191;
Best Local Similarity 19.78; Pred. No. 1.e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;

QY      21 YIELDLNGCELLESPREPERPMSTFKVLGAVLSRIDAGQELGRRIHYSONDLYEY 80
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      48 YCRSLNDSEY-----DMFGDYDSFTENSFIQVDDLEK-----YMQ----- 85

QY      81 SPYTEKHLTDGMTVRELCSAAITMSDNTANMLLTIGGPKELTAFLH----- 128
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      86 LPEHKHATDFAT-ENLCSESI-----KNKLSITITIGNLTELQTXKHTEHQSGYEGVTI 138

QY      129 -----NMGDHVTRLDRREPE-----LNEAIPND 151
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      139 EPGADLYDVPSSQAIFYFXNLQNSGNDLGDSMKDRDQXSSSHNTVNEELPHN 191

```

Search completed: September 10, 2003, 12:33:20
 Job time : 18.5714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds

(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_157T

Sequence: 1 HPEETLVKVKDAEDQLGARVGT.....TMDERNRQIAETIGASLIKHM 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:.*
2: p1r2:.*
3: p1r3:.*
4: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	534	38.9	105	2	JC2566
2	77	5.7	191	2	S67447
3	75.5	5.6	190	2	T09136
4	75.5	5.6	200	2	G97064
5	74.5	5.5	113	2	T45195
6	74.5	5.5	172	2	AD3606
7	73.5	5.5	195	2	T36975
8	72	5.3	149	2	F71252
9	71.5	5.3	152	2	T36984
10	71	5.3	153	2	E81708
11	71	5.3	177	2	A83753
12	71	5.3	192	2	G83096
13	70	5.2	108	2	E84217
14	70	5.2	170	2	AB0192
15	69.5	5.2	184	2	T21126
16	68.5	5.1	145	2	F84251
17	68.5	5.1	167	2	D87360
18	68.5	5.1	192	2	A83587
19	67.5	5.0	131	2	AD2281
20	66.5	4.9	116	2	C82906
21	66.5	4.9	177	2	D90227
22	66.5	4.9	180	2	C71869
23	66.5	4.9	198	2	D95285
24	65.5	4.9	42	2	F56978
25	65.5	4.9	181	2	AB1902
26	65.5	4.9	195	2	AE0623
27	65.5	4.9	198	2	C69296
28	65	4.8	146	2	C72703
29	65	4.8	148	2	E75283

30	65	4.8	150	2	T08585	calmodulin - soybe
31	65	4.8	160	1	E69186	conserved hypothet
32	65	4.8	168	2	B75498	conserved hypothet
33	65	4.8	177	1	B43387	nonstructural prot
34	65	4.8	177	2	J01931	nonstructural prot
35	65	4.8	180	2	G70912	hypothetical prote
36	64.5	4.8	145	2	H75262	hypothetical prote
37	64.5	4.8	151	2	D81333	probable protein-t
38	64	4.7	113	2	D70580	hypothetical prote
39	64	4.7	162	2	AG0769	probable protein-t
40	64	4.7	178	2	I40124	hypothetical prote
41	64	4.7	191	2	E95333	outer surface prot
42	63.5	4.7	164	2	H82336	hypothetical prote
43	63.5	4.7	166	2	C90029	regulator of sigma
44	63.5	4.7	168	2	T20606	hypothetical prote
45	63.5	4.7	179	2	AB1994	hypothetical prote

ALIGNMENTS

RESULT 1

JC2566

bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C:Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.

Gene 14811281, 81-86, 1994

A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MES>

C:Genetics:

A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match 38.9% Score 524; DB 2; Length 105;

Best local similarity 100.0%; Pred. No. 4.4e-37;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLRKLITGELLTSLRQOLIDMEADKVGPLRSALPAGFFIDKSGAGERSG 219

DB 2 AMATTLRKLITGELLTSLRQOLIDMEADKVGPLRSALPAGFFIDKSGAGERSG 61

QY 220 ITAALGPQKRSRIVYITTSQATMDERNRQIAETIGASLIKHM 263

DB 62 ITAALGPQKRSRIVYITTSQATMDERNRQIAETIGASLIKHM 105

RESULT 2

S67447

hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, March 1996

A:Reference number: 221766

A:Accession: T38062

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:Z69944; NID:g1217974; PIDD:CAA93808.1; PIDD:g1217978; GSPDB:C

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.7% Score 77; DB 2; Length 191;

Best Local Similarity 22.6%; Pred. No. 36;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
QY 104 MSDNTAANLLTTIGPKELTAFLHNKGDVTRLDNMEPLNEAIPNDEDDTTTPVAMAT 163
DB 1 MSVSHSLNLMQNPSCIGTIAILIVN-----VARID-----PASSKSTQOLVSMLN 46
QY 164 TIRKLTLGELLTLASRQQLIDMWEADKVAAPLLRSALPAGMTADKSGAGERSRGIIA 223
DB 47 EPRC-----IIRLPLGLYKLVNFRKSDSPETYSNAINIGITYE--GLAFLGKQIISI 99
QY 224 LGPDKP-----SNIVITYTGGQATMDERNROI----- 232
DB 100 -----SKPLEDKMLMSSRFMLDLTLITVQLREKTEDEKQHDLASNLASLPICITWS 155
QY 253 AEIGASLTKH 262
DB 156 VENGAGLHKH 165

RESULT 3
T09136
ADP-ribosylation factor homolog ARL3 - Trypanosoma brucei
N:Alternate names: ADP ribosylation factor 3 homolog
C:Species: Trypanosoma brucei
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T09136
R:Bringingud, F.; Vedrenne, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.; Mol. Biochem. Parasitol. 94, 249-264, 1998
A:Title: Conserved organization of genes in trypanosomatids.
A:Reference number: Z16580; MUID:98418771; PMID:9747975
A:Accession: T09136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190

A:Cross-references: EMBL:AF031926; NID:93452211; PIDN:AA032774.1; PID:93452215
A:Experimental source: strain AntAct1
A:Note: small G-protein
C:Genetics:
A:Gene: ARL3
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:89-94/Region: nucleotide-binding motif B
F:126-129/Region: GTP-binding NKXD motif
F:171-176/Region: nucleotide-binding motif B
F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

Query Match 5.6%; Score 75.5; DB 2; Length 190;
Best Local Similarity 20.5%; Pred. No. 48;
Matches 46; Conservative 37; Mismatches 78; Indels 63; Gaps 9;
QY 30 ELTSEFREPPEPMSTFKVLLCGAVLSRIDAGOEGLRRIHYSONDLYEVSPTYEKLT 89
DB 5 EELLKIRPPSR---RTRRIIMLGL---DNNGKRLRLRICEE-----VSDTRPT 48
QY 90 DGMTVEICSAITMSDNTAANLLTTIGPKELTAFLHNKGDV-----VTRL 137
DB 49 GGFNIGNITADEL-----KFFVWDVGGOKSLRSYRHYEDHDALVFVIDSADMERI 100
QY 138 DMEPELLEAIPNDEDDTTTPVAMATTLKLLTGELTLASRQQLIDMWEADKVAAPL-L 196
DB 101 EEARPELHILIEE-----KLGVVLLTLFANKODIPEAASOEYVSSINL 145
QY 197 RBAIPAGFIADKSG-AGERSGRGI-----AALGPDK 229
DB 146 ADTINRPMHIELSAETEGSLSSGWSVVDTLKKRRPRLRPQ 189
RESULT 4
G97064
spore coat protein COTJC [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Kolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200

A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:g15024270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

Query Match 5.6%; Score 75.5; DB 2; Length 200;
Best Local Similarity 25.7%; Pred. No. 51;
Matches 36; Conservative 26; Mismatches 55; Indels 23; Gaps 7;
QY 1 HPETLVKDAEDQGLARVGYIELDNLSEILSEFR-PEERPPMSTFKVLLCGAVLSRI 59
DB 10 HP---VKIKPNKQL-AKVITTYGGPDEGLASIVYLSQRSMTVP-----QATATLN 59
QY 60 DAQOEGLGRRIRHYSONDLYEVSPTYEKHLTDGTVRELCSAIT--MSDNTAANLLTTI 117
DB 60 DIGTEELAH-----LEIVGSIVQLSRGHSVELKSGIDAFADHDSATYPASAA 110
QY 118 GGPKELTAFILHNKGDVTRL 137
DB 111 GNP-FTAAVYIQSKGDPITDL 129

RESULT 5
T45195
hypothetical protein u1756t [imported] - Mycobacterium lepre
C:Species: Mycobacterium lepre
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z16911
A:Accession: T45195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

Query Match 5.5%; Score 74.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 28;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;
QY 130 MGDHVV---TRLDMEPELN---EAIPNDEDDTTTPVAMATTLKLLTGELLTLASRQ 180
DB 1 MGDGIGMEREGREKMGNTQCPLRVPGDESLIDGRASPEDLIT--NLSPITMSHPSP 58
QY 181 QLDWMEA-DKVAAPL-----RSALPAGFIADKSGAGERSRGIIAALGPDKPSR 232
DB 59 RDDDWVEPFDAIGTAVFADATGDKATMPAVAGI-----GASTRGS-GILASLSPFRPAR 112

RESULT 6
AD3606
molybdopteran biosynthesis mog protein [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AD3606
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
J. Mazur, M.; Gotsman, E.; Selkov, E.; Bizer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meli
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3606

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-172 <NUR>
A:Cross-References: GB:AE008918; PIDN:AAL54015.1; PID:q17984966; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0773
A:Map position: II
C:Superfamily: molybdenum cofactor biosynthesis protein B moab

Query Match
Best Local Similarity 25.5%; Score 74.5; DB 2; Length 172;
Matches 40; Conservative 14; Mismatches 42; Indels 61; Gaps 9;

QY 83 VTERKLTGDM-TYRELCSAATITMSDNTAANLLTTTG-GPRELTAFLHNMGDHYTRLDRLM 140
DB 39 IVRNRIPIPGMESVVR---TLIDLCDTVACDLITLTGGGPGS----- 76
QY 141 EPELNEAIPNDERDTTPVAMATTLRLKLT--GELTLASRQQLIDMEADKVAQPLLR 198
DB 77 -----PRDE-----TPKAKAVLHKELPFGFGOMRRVSLQ----- 107
QY 199 ALPGWFIADKSGAGSGRSGITIALGPDGKPSRIIV 235
DB 108 -TPYA-VLSRQTAGSRGKSF---LMLPGKPSIAM 138

RESULT 7

hypothetical protein SCJ11.04 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36975
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL data library, August 1999
A:Reference number: 221618
A:Accession: T36975
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-195 <OLI>
A:Cross-References: EMBL:AL109949; PIDN:CAB52869.1; GSPDB:GN00070; SCOEDB:SCJ11.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.04

Query Match
Best Local Similarity 17.0%; Score 73.5; DB 2; Length 195;
Matches 40; Conservative 43; Mismatches 77; Indels 75; Gaps 9;

QY 15 LGARVGTIEL-DLNGEILSFRRPFRPMSTFKVLLCGAVLSRIDAQEQGRRIHYS 73
DB 17 MASMLGILAEAREASAREVEVLREE-----AARAVALEAGEIELDRIV--- 60
QY 74 QNDLVESPYTERKHLTDGMTVRELCSAATITMSDNTAANLLTTTGPKELTAFHNMGDH 133
DB 61 -----TARELVETALVSAETGTEAGSEETALVPAASAAEPGAI----- 104
QY 134 VTLDLWREPELNEAT--PNDERDTTPVAMATTLRLKLTGELTLASRQQLIDMEADRV 191
DB 105 ---VPMOEGISVSLSPNNQR-----ILNVQDPRGLEPPRAKAI 142
QY 192 AGPLRSALPA-----GWFIDKSGAGSGRSGITIALGPDGKPS 231
DB 143 AALGTEAANAAYEGVPRPKRLAERGMILQESGAFSAGRR-LVAS--PGDDPS 194

RESULT 8

F71252
nucleoside-diphosphate kinase (EC 2.7.4.6) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 18-Jun-1999
C:Accession: F71252
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwilt

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: F71252
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-149 <COL>
A:Cross-References: GB:AE001269; GB:AE000520; NID:g3323336; PIDN:AAC65961.1; PID:g332
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP1010
C:Superfamily: nucleoside-diphosphate kinase
C:Keywords: ATP binding; hexamer; phosphohistidine; phosphoprotein; phosphotransferas
F;1418/Region: ATP binding #status predicted
F;117/Active site: His (phosphohistidine intermediate) #status predicted

Query Match
Best Local Similarity 22.2%; Score 72; DB 2; Length 149;
Matches 36; Conservative 24; Mismatches 64; Indels 38; Gaps 7;

QY 29 GEILSFRRPFRPMSTFKVLLCGAVLSRIDAGQOLGRRIHYSQNDLVESPYTERKL 88
DB 21 GEVLSEF---ERKGLVTLFALRLC-----YDPAFEL---HYAEHREKFPYSLIAYI 67
QY 89 TDGMTVRELCSAATITMSDNTAANLLTTTGKRELTAFLHNMGDHYTRIDNEPELNEAT 148
DB 68 TSAAPV-----ALAKGENALSLVRLTIGS-----TRVHAQP---GTI 103
QY 149 PND-ERDTTPVAMATTLRLKLTGELTLASRQQLIDMEAD 189
DB 104 RGDFALTRTNYIHVHSDSPESARRELALFVSQADPEWMDGN 145

RESULT 9

hypothetical protein SCJ11.13 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C:Accession: T36984
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL data library, August 1999
A:Reference number: 221618
A:Accession: T36984
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-152 <OLI>
A:Cross-References: EMBL:AL109949; PIDN:CAB52898.1; GSPDB:GN00070; SCOEDB:SCJ11.13
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.13
C:Superfamily: Streptomyces coelicolor hypothetical protein SCJ9.17

Query Match
Best Local Similarity 23.4%; Score 71.5; DB 2; Length 152;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

QY 106 DNFAANLLTTTIGKRELTAFLHNMGDHYTRIDNEPELNEAIPNDERDTTPVAMATTL 165
DB 25 DQAAARTIAQVTPPADLATFYESIGD-VT---WEDVGKGYFLNP----- 65
QY 166 RKLTLGELTLASRQQLIDMEADKVAQPLLRSAALPGWFIADKSGAGSGRSGITIALG 225
DB 66 ---AGDILLRLQEGVGVDFGTDEKSRGLV-----GSMGQ-GLIYVAG 104
QY 226 PDGKPSRIIVYITGSOATDERN-----ROIARI 255
DB 105 PDG-----AVYRTFT-ASLDEADLQVADGLRQPLEL 135

RESULT 10

EB1708

AB0192
conserved hypothetical protein YP01575 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AB0192
P:Parkhill, J., Wen, B.W., Thomson, N.R., Tibball, R.W., Holden, M.T.G., Prentice, M.B., deno-Tarragay, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Little, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrrell, N., Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-170 <RUB>
A:Cross-references: GB:AL590842; PIDD:G15979615; GSPDB:GN00175
C:Genetics:
A:Gene: YP01575
C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match	5.2%	Score 70;	DB 2;	Length 170;
Best Local Similarity	21.0%;	Pred. No. 1.2e+02;		
Matches	22;	Conservative	25;	Mismatches 44;
			Indels	14;
			Gaps	3

[illegible]

RESULT 15

C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
 C:Accession: T21126
 R:Steward, C.
 Submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19379
 A:Accession: T21126
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-184 <M1>
 A:Cross-references: EMBL:Z93378, PIDN:CA807583.1, GSPDB:GN00020, CESP:F19H8.3
 A:Experimental source: clone F19H8

C: Superfamily ADP-ribosylation factor
 K: Keywords: blocked amino end, lipoprotein, myristylation, nucleotide binding, P-loop
 E: 2-31/Region: nucleotide-binding motif A (P-loop)
 F: 90-95/Region: nucleotide-binding motif B
 F: 127-130/Region: GTP-binding NKXD motif
 F: 2/Modified site: myristylated amino end (gly) (in mature form) #status predicted

Query Match	5.2%;	Score 69.5;	DB 2;	Length 184;
Best Local Similarity	31.0%;	Score 71.0;	DB 1;	Length 184;

Matches	45;	Conservative	39;	Mismatches	81;	Indels	49;	Gaps	11.
---------	-----	--------------	-----	------------	-----	--------	-----	------	-----

```

QY      30 ELLESRRPERPPMSTFRVLLCGAVLSIDAGOEGLRHRYSONDIYVSPYTEKHILT 89
      ::::: | :::: | :::: | :::: | :::: |
Db      5 DYLSAKSKS----PSGREIRILLGL---DNAGKTTILKQL--SSEDVGHVTP-----T 48

```

DQ 90 DGGTVRELCSAATMSDNTAAANLLTTTIGPRELTAFLHMGDHV-----TRL 13
| : : | : || : : | ::
Db 49 KGFNVKTAA---MGD--IRLNWDIGGGRSIRPYWSNYENIDTLFVIDSNKKPF 10

QY 138 DRMEPELNEAIPNDERDITTPVAMATTLRLKLLTGELTLASROOLIDMEADKRVAGPIILR 19

Db 102 DENIELGELL-DEELTKRVPVILPILPANKO-----DLVTAASSEITRLNLD-----LLR 150

OY 198 SALPAGMTADSGAGERSRGITIALGPDGKPS 231

Db 151 DRT---WHIQASALKNECINDGIIYMWASNLKPA 181

Search completed: September 10, 2003, 12:26:18
Job time : 10.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds

(without alignments)
2219,902 Million cell updates/sec

Title: SEQ2_30E_157T
Perfect score: 1348
Sequence: 1 HPETLVKVNADPDIGARVG.....TMDERNQIAETGASLIRKW 263

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	77	5.7	191	1	YDA4_SCHPO
2	75	5.6	201	1	RAC2_DICDI
3	72	5.3	149	1	NDK_TREPA
4	68.5	5.1	196	1	RAC2_LOTJA
5	67	5.0	122	1	HC3L_TRIFE
6	66.5	4.9	116	1	RBP4_UREPA
7	65.5	4.9	198	1	OGG1_ARCFU
8	65.5	4.9	206	1	KTRH_METAC
9	65	4.8	177	1	VNSC_RINDR
10	65	4.8	177	1	VNSC_RINDR
11	63.5	4.7	144	1	MARR_ECOLI
12	63.5	4.7	185	1	YCJC_ECOLI
13	63.5	4.7	196	1	ALKH_BACSU
14	63	4.7	176	1	HSUV_THEMA
15	62.5	4.6	150	1	PDOV_SALTY
16	62.5	4.6	172	1	YDEL_ECOLI
17	62.5	4.6	182	1	PYRE_SYRGO
18	62.5	4.6	195	1	TRPE_THEVO
19	62.5	4.6	200	1	RR4_PELNE
20	62	4.6	174	1	THAB_RAT
21	62	4.6	182	1	YCEB_SALTY
22	62	4.6	186	1	YCEB_SALTY
23	62	4.6	202	1	COAT_ELV
24	61.5	4.6	178	1	HSUV_RALSO
25	61	4.5	121	1	SECR_HUMAN
26	61	4.5	200	1	TARF_CAUCR
27	60.5	4.5	103	1	RS10_NEIGO
28	60.5	4.5	146	1	HBG_RABIT
29	60.5	4.5	156	1	BFR_AZOVI
30	60.5	4.5	160	1	TATB_HELPY
31	60.5	4.5	175	1	Y581_AQUAE
32	60.5	4.5	184	1	MLR1_SCHPO
33	60	4.5	148	1	CALM_BLAEM

34	60	4.5	159	1	GRE4_CHLME
35	60	4.5	178	1	PYRE_ARCFU
36	60	4.5	178	1	UCRI_ANKSP
37	60	4.5	184	1	ARL2_DROME
38	59.5	4.4	103	1	RS10_NEIMA
39	59.5	4.4	135	1	CCRN_PAROL
40	59.5	4.4	152	1	YU3_YERPE
41	59.5	4.4	184	1	HRPL_PSESE
42	59.5	4.4	196	1	1196_ASPB7
43	59.5	4.4	198	1	UPP_AKATH
44	59.5	4.4	205	1	RS4_ECOLI
45	59.5	4.4	206	1	RS4_SHRPL

ALIGNMENTS

RESULT 1
YDA4_SCHPO
ID YDA4_SCHPO STANDARD; PRT; 191 AA.
AC Q10346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F12.04c in chromosome I.
GN SPAC1F12.04c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP
RC
RX MEDLINE=21848401; PubMed=11859360;
SR STRAIN=972;
SEQUENCE FROM N.A.
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris N., Hildes J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Egger P., Zimmermann W., Medler H., Manbut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez J., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Usery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe".
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; 269944; CAAG3808.1; -
DR PIR; T38062; S67447.
DR GeneDB_Spombe; SPAC1F12.04c; -

```

KW Hypothetical protein.
SQ SEQUENCE 191 AA; 21549 MW; 65555347E0EBED16 CRC64;

Query Match 5.7%: Score 77; DB 1; Length 191;
Best Local Similarity 22.6%; Pred. No. 26;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;

QY 104 MSDVTANILDTTIGGPELPAFLNNQGDYTRLDREMEPELNEAIPNDEROTTPVAMAT 163
DB 1 MSYHSHLMLQNSGIDKIALIYN---VARD-----PASKSTAQLYSMLN 46
QY 164 TLRLKLTGELLTLASROOLIDMWEADKACPLBSALPAGFIADKSGAGERSGRIIAA 223
DB 47 EFRG-----ILRLGLYKLIYNFRDSDSEPTYSNAINIGYYTE--GLAELGKQIISI 99
QY 224 LGPGGRK-----SRIVYITTGSAQNDERROI----- 252
DB 100 ---SKPLEDKLWMSRFWLLDTLTLTYQLIREKTEDEKHOOLDLASNLASLPICHW 155
QY 253 AEIGASGLIKH 262
DB 156 VENGAGLHKH 165

RESULT 2
RACG_DICDI STANDARD; PRT; 201 AA.
ID RACG_DICDI STANDARD; PRT; 201 AA.
AC Q9GGS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RAS-related protein racg.
GN RACG.
OS Dictyostellum discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=21127961; PubMed=11222756;
RA Rivero F., Dielisch H., Glockner G., Noegel A.A.;
RL "The dictyostellum discoideum family of rho-related proteins.";
Nucleic Acids Res. 29:1068-1079(2001)
RL -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; AF310893; AAG45130.1; -.
DR HSSP; P21181; 1AM4.
DR DICTYDB; DD27772; racg.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR001806; Ras_Lnrmstmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASSTRNSFRMNG.
DR SMART; SM00174; RHO; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 57 61 GTP (BY SIMILARITY).
FT NP_BIND 115 118 GTP (BY SIMILARITY).
FT DOMAIN 32 40 EFECTOR REGION (POTENTIAL).
FT LIPID 198 198 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 201 AA; 22813 MW; F6C5B86497CCCL69 CRC64;

Query Match 5.6%; Score 75; DB 1; Length 201;
Best Local Similarity 22.6%; Pred. No. 40;

```

	Matches	36;	Conservative	20;	Mismatches	59;	Indels	44;	Gaps	7;
QY	44	MSTRFVLICGAVLSRIDAGQEQGLRRHRYSONDLV-EYSP-VYEKHLTDGNTVRELCSAA	101							
Db	1	MKSIVCVYVG-----EGGIGKTSMLSYSNSISNEYOPVFDNYST-----	42							
QY	102	ITMSDNTNAANLLTTTIGSPKRELFATP-----LHMGCHVRLDRWEPELN	145							
Db	43	LHMNKKRPYNLSLMDTGAQEEFSLRLRLSYQTVFLLCSLINPSSFNSLDSVQDELN	102							
QY	146	EAIPINDERDTTPVAMAATTLEKLLTIGELL-TLASROOL	182							
Db	103	ENCEN-----TPIVLVGTQMDLKSNSVYILDRICEKKQL	135							
RESULT 3										
ID	NDK_TREPA	STANDARD:	PRT:	149	AA.					
AC	083974:									
DT	15-DEC-1998	(Rel. 37, Created)								
DT	15-DEC-1998	(Rel. 37, Last sequence update)								
DT	28-FEB-2003	(Rel. 41, Last annotation update)								
DE	Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)									
DE	(Nucleoside-2-P kinase).									
GN	NDK OR TP1010.									
OS	Treponema pallidum.									
OC	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.									
OX	NCBI_TaxId=160;									
RM	[1]									
RP	SEQUENCE FROM N.A.									
RC	STRAIN-Nichols;									
RA	MEDLINE:98332770; PubMed=9665876;									
RA	Fraser C.M., Norris S.J., Welstock G.M., White O., Sutton G.G.,									
RA	Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,									
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,									
RA	Khakhal H., Richardson D., Howell J.K., Chidambaram M., Utechtack T.,									
RA	McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,									
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,									
RA	Venter J.C.;									
RT	*Complete genome sequence of Treponema pallidum, the syphilis									
RT	spirochete.;									
RL	Science 281:375-388(1998).									
CC	-1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES									
CC	OTHER THAN ATP.									
CC	-1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate -> ADP +									
CC	nucleoside triphosphate.									
CC	-1- SUBUNIT: Homotetramer (By similarity).									
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).									
CC	-1- SIMILARITY: Belongs to the NDK family.									

```

RESULT 4
RAC2 LOTJA
ID RAC2 LOTJA STANDARD; PRT; 196 AA.
AC 040220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAC-1like GTP binding protein RAC2.
GN RAC2.
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Finales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
OX NBIL_TaxID=34303;
XX (1)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Gifu / B-129; TISSUE=Root nodules;
RX MEDLINE=97231679; PubMed=9076991;
RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
RT Identification of new protein species among Lotus japonicus GTP-
RT binding proteins encoded by cDNAs from Lotus japonicus, and
RT expression of corresponding mRNAs in developing root nodules.*;
RL Plant J. 11:237-250(1997).
CC - SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

Db      1  MSTARFIKC-----VTYODGAVGKTCMLISTSTNTPPDYVP-----TVDFNFS 44
QY      100  AATMSDNTANANILLTTTGGPRE-----LTAFL-LHNMGDHYTRLDRMEPE 144
Db      45  ANYVV- -DGSVTYNLGLMDTGAQEDYNNRLRPLSYRGADYFLNAFLSLLSRASYNNISKRIPE 104
QY      144  LNEAINDERDHTTPYAAATTLRKLLITGELLTL-ASRQOLIDMWEADKVA---GPLRISA 199
Db      104  LRHYAP-----TVPI-----VLVGTKLDLRDROLYLIDHPGATPTTAAQSEELKKA 149
QY      200  LPAGAFI 206
Db      150  IGAAYTL 156

```

ID	HC3L_THIPE	STANDARD;	PRT;	122 AA.
AC	P80509;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	Cytochrome-c3 hydroxylase, large chain (EC 1.12.2.1) (Hydroxylase)			
DE	(Fragments).			
GN	HOXG.			
OS	Thiobacillus ferrooxidans.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;			
OC	Acidithiobacillaceae; Acidithiobacillus.			
OX	NCBI_TaxID=920;			
RN	[1]			
RP	SEQUENCE.			
RX	STRAIN=ATCC 19859;			
RX	MEDLINE=96241862; PubMed=8661919;			
RA	Fischer J., Quentmeyer A., Kostka S., Kraft R., Friedrich C.G.;			
RT	Purification and characterization of the hydroxylase from			
RT	Thiobacillus ferrooxidans."			
RL	Arch. Microbiol. 165:289-296(1996).			
CC	-1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 + 2			
CC	ferrocitochrome c3.			
CC	-1- COFACTOR: IRON.			
CC	-1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS			
CC	ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.			
KW	Oxidoreductase; Iron.			
FT	NON_CONS	20	21	
FT	NON_CONS	29	30	
FT	NON_CONS	35	36	
FT	NON_CONS	42	43	
FT	NON_CONS	59	60	
FT	NON_CONS	72	73	
FT	NON_CONS	78	79	
FT	NON_CONS	87	88	
FT	NON_CONS	98	99	
FT	NON_CONS	107	108	
FT	NON_TER	122	122	
ST	SEQUENCE	122 AA;	EB4FAF36534BDA0B	CRC64;
Query Match	5.0%;	Score 67;	DB 1;	Length 122;
Best Local Similarity	26.4%;	Pred. No. 93;		
Matches	28;	Conservative 13;	Mismatches 23;	Indels 42;
				Gaps 4;
QY	19 VGYLELDINSEILLESFPERPEPMSTFKVLLCGAVLSRIDAGCEQIGRRIRHSQNDLV 78			
DB	12 VGRVEGDSDXSILEYFR-----NAILARRGGG-----LG 41			
QY	79 EYSPVTEKHLDGATVRELCSAATMSDNTANLLTTIGSKPRELT 124			
DB	42 KVAFTGTGNYEIGVT-----ISGDKDPQAGLVVT---PREST 75			

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR U0321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3.
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.G., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caesell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 50S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002129; AAF30730.1; -
DR HAMAP; MF_00003; -; 1.
DR InterPro; IPR000238; R1b_bind_facctA.
DR Pfam; PF02033; RBFA; 1.
DR ProDom; PD007327; R1b_bind_facctA; 1.
DR TIGRFAMs; TIGR00082; rbfA; 1.
DR PROSITE; PS01319; RBFA; FALSE NEG.
KW rRNA processing; Complete proteome.
SQ
SEQUENCE 116 AA; 13247 MW; A79DCC1F0547514 CRC64;

Query Match
Best Local Similarity 4.9%; Score 66.5; DB 1; Length 116;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

OY 144 LNEALPDERDRTTPAAMTLKLTGL-----LTLASROQLDME-ADKVAGPLL 196
DB 18 INNALANEINDKIATLARTAVR--LSNDLSVAKIFLDHKKESMLKLVNKKVSG-LL 74

OY 197 RSALPAGM-----FIADKS 210
DB 75 RSKLLAEKTSYKVPRLRFYIDR 97

RESULT 7
OGGI_ARCFU STANDARD; PRT; 198 AA.
AC 029876;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN OGG OR AF0371.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49556;

```

```

RX MEDLINE=98049343; PubMed=9389475;
RX Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson R.E.,
RX Ketchum K.A., Dodson R.J., Gwim M., Hickey E.K., Peterson J.D.,
RX Richardson D.L., Kerlavage A.R., Graham D.E., Kyriides N.C.,
RX Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RX Riess E.F., Dougherty B.A., McKenney R., Adams M.D., Loftis B.,
RX Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodek A., Zhou L.,
RX Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RX Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RX Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RX Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RX Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
CC guanine (7,8-dihydro-8-oxoguanine = 7-oxoG) from DNA. Also nicks
CC DNA at apurinic/apyrimidinic sites (AP sites) (by similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001079; AAB90876.1; -
DR PIR; C69296; C69296.
DR TIGR; AF0371; -
DR HAMAP; MF_00241; -; 1.
DR InterPro; IPR003265; Endo_3c.
DR Pfam; PF00730; HHN-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR Hydrolase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
KW Multifunctional enzyme; Complete proteome.
FT ACT SITE 122
FT ACT SITE 122
FT SEQUENCE 198 AA; 22639 MW; 3A5C033AA12F3FB CRC64;

Query Match
Best Local Similarity 4.9%; Score 65.5; DB 1; Length 198;
Matches 36; Conservative 19; Mismatches 54; Indels 35; Gaps 7;

OY 14 OLGANVGITIEDINSGETLSEFRPEPRPMSTFKVLLCGAVL---SRIDAG---OEOLG 67
DB 15 QLGER-----GEVEFDFRPLDPSVATITRTELAFISTRANSSATGGLKFORLHG 64

OY 68 RRIHYSQ-----NDLVEYSPYTERKHLTDGNTVRELCSAAITMSDNTAANLLTTI 117
DB 65 QGVGVKEALLTAGVFRHNRKAEY-----IREAFKSFLEVKALESSKAREILLKIK 117

OY 118 G-GPKELTAPFLHNG-DHYTRIDR 139
DB 118 GIGMKRASHFLRNVRGDEVAIIDR 141

RESULT 8
KTHY_METAC STANDARD; PRT; 206 AA.
AC 087HS9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR MA4433.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders Incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.

```



```

OX NCBI_TaxID=2214;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas D.E., Graham D.A., Guss A.M.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Ye W.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zander S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosaerina acetivorans reveals extensive metabolic
RL and physiological diversity."
RL Genome Res. 12:532-542(2002).
CC
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: AE011164; AM07774.1; -.
DR HAMAB: MF_00165; -. 1.
DR InterPro: IPR000662; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR TIGRPFAMs: TIGR00041; DTPM_Kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
DR TRANSFERase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C2E5BC85423B3 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.7%; Pred. No. 2.4e+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

OY 56 LSRID-AGGEOGLRRHYGNDVEXSPTEKHLTGKTVRELCSAATMSONTAANLL 114
DB 8 LSGIDSGSKSTVAKKIQ-ENSELRVEPEVTRPRETRG-----TLTGNAYENAIQ 55
OY 115 TTIGPKELTAFLHNMGDVTRLDREPELNEAIPDERDOTTVPAMATTIRKLTLGELL 174
DB 56 SDPTQALAEFLFTADIAEHLAKYKPALEDGKTVISDRSKAYATGGITLKRRLDNPJ- 114
OY 175 TLASRQQLDWMEDAKVAGPLRLSALPAGW-----FTAD-----KSG-AGE----- 214
DB 115 -----EWVR-----DHRGWTVPIDLTFLIEDIEPEIAVARCGRGQTTF 154
OY 215 -----RGSRGITAAIGPDGKPSRIYVITTSQATMDE 247
DB 155 EKIEPLRGVRELFLGLAE-EPERFYVDASGSPEDVER 192

RESULT 9
VNSC_RINDK STANDARD: PRT: 177 AA.
AC P35948;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein C.
CC
CC Rinderpest virus (strain Kabete O) (RDV).
OS

VIRUSES: ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID=11242;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92410646; PubMed=1529555;
RA Yamanaka M., Dale B., Crisp T., Cordell B., Grubman M., Ylma T.;
RT "Sequence analysis and editing of the phosphoprotein (P) gene of
RL rinderpest virus."
RL Virology 190:553-556(1992).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: S44819; AB23269.1; -.
DR PIR: B43387; B43387.
DR InterPro: IPR003875; Paramyxovir_NS_C.
DR Pfam: PF02725; Paramyxo_NS_C; 1.
KW Nonstructural protein.
SQ SEQUENCE 177 AA; 19926 MW; 116971B140A39F11 CRC64;

Query Match 4.8%; Score 65; DB 1; Length 177;
Best Local Similarity 24.6%; Pred. No. 2.2e+02;
Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

OY 93 TVRELCSAATMSDNTAANLLTTIGPKELTAFLHNMGDVTRLDREPELNEAIPNE 152
DB 49 TIRISASHSQDQAKKACIATVINDLEATVMS-----WEHSL----- 90
OY 153 RDTTTP-----YAAATIRKLTLGELLTLASRQQLDWMEDAKVAGPLRLSALPA 202
DB 91 ----VTQCCAPRYSITFMITAVKRLRESKMLTSLWFNOL--MAYSK-SGGEEMRLRTA 144
OY 203 GWEFIAD 208
DB 145 MMTLAN 150

RESULT 10
VNSC_RINDK STANDARD: PRT: 177 AA.
AC Q03339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Nonstructural protein C.
GN C.
OS Rinderpest virus (strain RBOK) (RDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID=36409;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93155658; PubMed=8429304;
RA Baron M.D., Shalla M.S., Barrett T.;
RT "Cloning and sequence analysis of the phosphoprotein gene of
RL J. Gen. Virol. 74:299-304(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: X68311; CA448391.1; -.

```


ID YCJC_ECOLI STANDARD; PRT; 185 AA.
AC P38522; P76839; P77417;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ycjC.
GN YCJC OR B1299 OR SFL304.
OS Escherichia coli, and
OC Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-E.coli; STRAIN-K12 / MG1655;
RX MEDLINE-97426517; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1233-1238(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-E.coli; STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Ogihara T., Saito N.,
RA Sampaio G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 65-185 FROM N.A.
RC SPECIES-E.coli;
RX MEDLINE-91216440; PubMed-1840553;
RA Helm R., Strehler E.R.;
RT "Cloning an Escherichia coli gene encoding a protein remarkably
similar to mammalian aldehyde dehydrogenases.";
RL gene 99:15-23(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-S.flexneri; STRAIN-301 / Serotype 2a;
RX MEDLINE-22272406; PubMed-12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: Insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [5]
RP IDENTIFICATION.
RC SPECIES-E.coli;
RX Rued K.E.;
RT Unpublished observations (AUG-1994).
RN [6]
RP -1- SIMILARITY: SOME, TO E.INFLUENZAE H10659.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE000228; AAC74381.1; -
CC EMBL; D90768; BAA14868.1; -
CC EMBL; D90767; BAA14859.1; -

DR EMBL; M38433; -. NOT_ANNOTATED_CDS.
DR EMBL; AE015157; BAA42915.1; -
DR PIR; F64878; F64878.
DR EcoGene; EC12431; ycjC.
DR InterPro; IPR007113; Cupin_sup.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3.1.
DR SMART; SM00530; HTH_XRE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 185 AA; 20091 MW; D7DD3FD794A7768 CRC64;
Query Match 4.7%; Score 63.5; DB 1; Length 185;
Best Local Similarity 27.5%; Pred. No. 3e+02;
Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;
OY 145 NEAIPNDERDTPVPAATTLKRLTGLTLTASROQLIDWMEADKVGAPLRSALPAGW 204
DB 34 HSAISTIEQDKVSPA--ISTQLKLVGLSLSE-----FSEPERPDEPOV----- 78
OY 205 FIADKSGAERSRSGRIALGPDGKRSRLVTV-----YTTGSGATMDR-NQIAEIGASL 259
DB 79 -VIMDDLIEMSGSGVSMKLVHNGNPNRTIAMIFETYPGT--TTGERIKHGGEIGVYL 135
RESULT 13
ALRX_BACSU STANDARD; PRT; 196 AA.
AC P50846;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE KHG/KDPG aldolase [Includes: 4-hydroxy-2-oxoglutarate aldolase
DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KHG-aldolase); 2-
DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
DE dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate
DE aldolase)] (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
DE aldolase).
GN KDCGA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / Marburg;
RX MEDLINE-96349105; PubMed-8760912;
RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
the *serA* and *kdg* loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borbely S.,
RA Borries R., Boursier L., Brans A., Brann M., Brinell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Etlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizi E., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Potwolk S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.-J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambitt R., Wedler H., Wedler H., Wetzenecker T.,
 RA Winters P., Wipert A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate + pyruvate +
 CC glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate -
 CC pyruvate + D-glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GYVOXYLATE.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KMG/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L47838; AAB38480.1; -;
 DR EMBL: 299115; CAB1427.1; -;
 DR PIR: H69647; H69647.
 DR Subtilast: BG11386; kdgA.
 DR InterPro: IPR000887; Aldase_KDPG_KMG.
 DR Pfam: PF01081; Aldolase_1.
 DR TIGRFAMS: TIGR01182; eda; 1.
 DR PROSITE: PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
 DR PROSITE: PS00160; ALDOLASE_KDPG_KMG_2; 1.
 KM Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
 FT ACT_SITE 43 43 BY SIMILARITY.
 FT ACT_SITE 47 47 BY SIMILARITY.
 FT BINDING 130 130 SCHIFF-BASE WITH KMG OR PYRUVATE (BY
 FT SIMILARITY).
 SQ SEQUENCE 196 AA; 20865 MW; EB0B597FD1102CC4 CRC64;
 Query Match 4.7%; Score 63.5; DB 1; Length 196;
 Best Local Similarity 21.6%; Pred. No. 3.2e+02;
 Matches 27; Conservative 18; Mismatches 61; Indels 19; Gaps 4;
 OY 28 SGEIESFRPEFRPMSTFKVLICAVLSRIDAGQ-EQLGRRIHSON--DIYVSPV 83
 DB 51 ASDIIESFRNRREDI-----LIGAGYISAOAGAEAKAGAOIFVSGFSADLAHLSF 103
 OY 84 TEKHLTDGWTYRELCSAITMSDNTAANL-----LTTIGCKEELTAFLHNGDHYT 135
 DB 104 VTHYIPGVLPSEIMEALTFGFTTLKLPSPGVGFCIPMKNLGFPQVTEIPGCIHPS 163
 OY 136 RLDKW 140
 DB 164 EYPDW 168
 RESULT 14
 HSLV_THEME STANDARD; PRT; 176 AA.
 ID HSLV_THEME
 AC Q9WZ1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-dependent protease hslv (EC 3.4.25.-).
 GN HSLV OR TM0521.

OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MS8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.R., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of *Thermotoga maritima*.
 RL Nature 399:323-329(1999).
 CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (By similarity).
 CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001728; AAD35606.1; -;
 DR PIR: G72365; G72365.
 DR HSSP: P31059; 1E94.
 DR MEROPS: T01.006; -.
 DR TIGR: TM0521; -.
 DR HAMAP: MF_00248; -; 1.
 DR InterPro: IPR001353; Protease_protease.
 DR Pfam: PF00227; Proteasome_1.
 KM Hydrolase; Protease; Complete proteome.
 FT ACT_SITE 6 6 BY SIMILARITY.
 SQ SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;
 Query Match 4.7%; Score 63; DB 1; Length 176;
 Best Local Similarity 28.2%; Pred. No. 3.1e+02;
 Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;
 OY 117 ICGPKELTAFLHNGDHYTRIDRWPELEAIPNDERDTTPVAMATTLKLLTGELTL 176
 DB 42 LGEGKVLGFRAGSVADAMTLFDRPEAKRE-----WCGNLTK 78
 OY 177 ASROQLDMKEDAKVAGPLRSALPGWFTADR-----SGAGERSRGII-----AAL 224
 DB 79 AAVELADW-RTRV--LR-RLEALLVADRENIFIIISNGE-----VTPQDDAAAI 127
 OY 225 GPDG 228
 DB 128 GSGG 131
 RESULT 15
 PDUV_SALTY STANDARD; PRT; 150 AA.
 ID PDUV_SALTY
 AC Q9XDM6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Propionediol utilization protein pduv.
 GN PDUV OR STM2056.
 OS Salmoneella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmoneella.
 OX NCBI_TaxID=602;
 RN (1)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cells updates/sec

Title: SEQ2_30E_157T
Perfect score: 1348
Sequence: 1 HPERLVKXKADQDQAGAVG.....TMDERKQIAIGASLIRHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHES:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	524	38.9	105	2	Q52026	pseudomonas
2	524	38.9	105	2	Q52330	escherichia
3	499	37.0	145	2	Q8RTD8	klebsiella
4	488	36.2	102	2	Q52639	pseudomonas
5	464	34.4	138	2	Q990F3	escherichia
6	464	34.4	138	2	Q9AMM1	escherichia
7	464	34.4	139	2	Q9AMA2	escherichia
8	462	34.3	138	2	Q9AMAO	escherichia
9	462	34.3	138	2	Q9AM99	escherichia
10	462	34.3	139	2	Q9AM98	escherichia
11	459	34.1	139	2	Q9AMM3	escherichia
12	338	23.1	67	2	Q53553	shigella fl
13	326	24.2	95	2	Q8GDES	klebsiella
14	310	23.0	62	2	Q9JN58	shigella fl
15	301	22.3	128	2	Q8VQ00	escherichia
16	251	18.6	180	2	Q8KVT2	staphylococ

17	246	18.2	52	2	Q9R412	shigella fl
18	241	17.9	48	2	Q9RLH0	proteus mir
19	197	14.6	38	2	P97145	escherichia
20	164.5	12.2	134	2	Q8VUL3	staphylococ
21	119.5	8.9	109	2	Q53698	staphylococ
22	105	7.8	20	2	P97146	escherichia
23	87	6.5	100	2	Q93S05	staphylococ
24	82	6.1	68	2	Q9XBJ2	staphylococ
25	80	5.9	192	2	Q9ACM8	staphylococ
26	79.5	5.9	108	2	Q985Z0	staphylococ
27	75.5	5.6	190	5	Q76Z27	trypanosoma
28	75.5	5.6	200	16	Q97JF8	clostridium
29	75	5.6	181	16	Q8D108	yersinia pe
30	74.5	5.5	113	2	Q49970	mycobacteri
31	74.5	5.5	172	16	Q8YBW3	bruceella me
32	74.5	5.5	172	16	Q8YWF9	bruceella su
33	73.5	5.5	195	16	Q9RIAO	streptomyce
34	73.5	5.5	204	11	Q91YH4	mus musculi
35	73	5.4	131	2	Q9XSH0	yersinia en
36	73	5.4	153	5	Q8S5Z0	citroni intes
37	73	5.4	175	17	Q8ZTH8	pyrobaculum
38	72.5	5.4	145	2	Q05984	staphylococ
39	72	5.3	205	16	Q98JF8	rhizobium l
40	71.5	5.3	152	16	Q9R191	streptomyce
41	71.5	5.3	196	2	Q9AH34	pseudomonas
42	71	5.3	150	10	Q94IG4	nicotiana t
43	71	5.3	153	16	Q9PKT4	chlamydia m
44	71	5.3	177	16	Q9KEM7	bacillus ha
45	71	5.3	192	16	Q9HM07	pseudomonas

ALIGNMENTS

RESULT 1

Q52026	PRELIMINARY;	PRT;	105 AA.
ID	Q52026;		
AC	Q52026;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Beta-lactamase (Fragment).		
GN	BLA.		
OS	Pseudomonas aeruginosa.		
OC	Plasmid PR01614.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
OC	Pseudomonadaceae; Pseudomonas.		
OX	NCBI_TaxID=287;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-95011664; PubMed-7926843;		
RA	West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;		
RT	*Construction of improved Escherichia-Pseudomonas shuttle vectors		
RT	derived from pUC18/19 and sequence of the region required for their		
RT	replication in Pseudomonas aeruginosa.*;		
RL	Gene 148:81-86(1994).		
DR	EMBL; L30112; AAA66058.1; -		
DR	HSSP; P00810; 1XPB.		
DR	InterPro; IPR001466; Beta_lactamase.		
DR	InterPro; IPR000871; Beta_lactamase_A.		
DR	Protein; P000144; beta-lactamase; 1.		
DR	PRINTS; PR00118; BLACTAMASEA.		
KW	Plasmid.		
FT	NON_TER		
SO	SEQUENCE	105 AA;	11229 MW; D28899A073330557 CRC64;
QY	Query Match	38.9%;	Score 524; DB 2; Length 105;
QY	Best Local Similarity	100.0%;	Pred. No. 1e-36;
QY	Matches 104;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
DB	160 ANAATLRKLTGELLTLASROOLIDMEADKVGPIILRSALPRGWIADKSGGSGSG 219		
DB	2 ANAATLRKLTGELLTLASROOLIDMEADKVGPIILRSALPRGWIADKSGGSGSG 61		

QY 220 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 263
 Db 62 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 105

RESULT 2

Q52330 PRELIMINARY; PRT; 105 AA.
 ID 052330
 AC 052330;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Tnl bla protein (Fragment).
 DE Escherichia coli.
 OS plasmid RK2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NC NCBL_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90264294; PubMed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The k11-kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korC.";
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL; M32794; AAA26408.1; -.
 DR HSSP; P00810; 1XPB.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS; PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 38.9%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1e-36;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 219
 Db 2 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 61
 QY 220 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 263
 Db 62 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 105

RESULT 3

Q52330 PRELIMINARY; PRT; 145 AA.
 ID 052330
 AC 052330;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 DE Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 NC NCBL_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palasubramaniam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT Subtilactone (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER
 SQ SEQUENCE 145 AA; 145 MW; D2889A4073330557 CRC64;

Q52330 PRELIMINARY; PRT; 145 AA.
 ID 052330
 AC 052330;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 DE Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 NC NCBL_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palasubramaniam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT Subtilactone (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER
 SQ SEQUENCE 145 AA; 145 MW; D2889A4073330557 CRC64;

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;
 Query Match 37.0%; Score 499; DB 2; Length 145;
 Best Local Similarity 68.8%; Pred. No. 2.1e-34;
 Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 100 AATMSDNTAANLLTTIGGPKELTAFIANNKGDHVTFLDWEPELNAIPNDERDTTPY 159
 Db 7 AAYMSDNSAANLLTATVGGPAGLTAFIRQIGDNVTRLDRETELNEALPGDARDTTPA 66
 QY 160 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 219
 Db 67 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 126

QY 220 IIAALGPDGKPSRIIVY 237
 Db 127 IVALGPNNAERIVY 144

RESULT 4

Q52639 PRELIMINARY; PRT; 102 AA.
 ID 052639
 AC 052639;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Beta-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC plasmid PRO1600.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NC NCBL_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94336757; PubMed=8058819;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the pOR
 RT PRO1600 Oritr region found in the broad-host-range plasmids of the pOR
 RT series.";
 RL plasmid 31:265-274(1994).
 DR EMBL; L22691; AAA96312.1; -.
 DR HSSP; P00810; 1XPB.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS; PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER
 SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.2%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1.1e-33;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 219
 Db 2 AMATTLLKLTGELLTASRQQLIDWMEADKVAAGPLIRSLPAGWFIADKSGAGERSRG 61
 QY 220 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 263
 Db 62 IIAALGPDGKPSRIIVYITGSOATMDERNRQIAETGASLIIKH 105

RESULT 5

Q52639 PRELIMINARY; PRT; 138 AA.
 ID 052639
 AC 052639;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327049; AK07464.1; -;
DR EMBL: AF327051; AK07466.1; -;
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYRLDRWEPELNEAIPNDRDRTTTPYAAATTLTKLTT 170
DB 10 NLLATVGGPAGLTAFRLRGIDNVTRLDRWETELNEALPGDARDTTTPASMAATLTKLTT 69
OY 171 GELLTLASROQLIDWNEADKVAGPLLRSLAPGWFADKSGGERSGRTIIAALPGDGRP 230
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKTGASKRGARIVALLGPNNA 129
OY 231 SRIVVIY 237
DB 130 ERIVVIY 136

RESULT 6
O9AMA1 PRELIMINARY; PRT; 138 AA.

AC O9AMA1:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1B/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327050; AK07465.1; -;
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYRLDRWEPELNEAIPNDRDRTTTPYAAATTLTKLTT 170
DB 10 NLLATVGGPAGLTAFRLRGIDNVTRLDRWETELNEALPGDARDTTTPASMAATLTKLTT 69
OY 171 GELLTLASROQLIDWNEADKVAGPLLRSLAPGWFADKSGGERSGRTIIAALPGDGRP 230
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKTGASKRGARIVALLGPNNA 129
OY 231 SRIVVIY 237
DB 130 ERIVVIY 136

RESULT 7

O9AMA2 PRELIMINARY; PRT; 139 AA.
AC O9AMA2:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT Isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327048; AK07463.1; -;
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 139;
Best Local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYRLDRWEPELNEAIPNDRDRTTTPYAAATTLTKLTT 170
DB 10 NLLATVGGPAGLTAFRLRGIDNVTRLDRWETELNEALPGDARDTTTPASMAATLTKLTT 69
OY 171 GELLTLASROQLIDWNEADKVAGPLLRSLAPGWFADKSGGERSGRTIIAALPGDGRP 230
DB 70 SQRLSARSGROLQWVDDRVAGPLIRSVLPAGWFIADKTGASKRGARIVALLGPNNA 129
OY 231 SRIVVIY 237
DB 130 ERIVVIY 136

RESULT 8

O9AMA0 PRELIMINARY; PRT; 138 AA.
AC O9AMA0:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-E/98 9-1:
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia."
DR EMBL: AF327052; AAK07467.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLTLT 170
DB 10 NLLATYGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDARTTTPASMAATLRLTLT 69
QY 171 GELLITLARSQQLIDMEADKVAAGPLLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
DB 70 SQRLSARSQRLQGMVDDRVAGPLIRSVLPAGWFIADKSGAGSGRGITIALGPNKKA 129
QY 231 SRIIVYI 237
DB 130 ERIIVLY 136

RESULT 9
Q9AM99 PRELIMINARY; PRT; 138 AA.
ID Q9AM99;
AC Q9AM99;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-E/99 4-1;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia."
DR EMBL: AF327053; AAK07468.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLTLT 170
DB 10 NLLATYGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDARTTTPASMAATLRLTLT 69
QY 171 GELLITLARSQQLIDMEADKVAAGPLLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
DB 70 SQRLSARSQRLQGMVDDRVAGPLIRSVLPAGWFIADKSGAGSGRGITIALGPNKKA 129

```

```

QY 231 SRIIVYI 237
DB 130 ERIIVLY 136

RESULT 10
Q9AM98 PRELIMINARY; PRT; 139 AA.
ID Q9AM98;
AC Q9AM98;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-E/99 3-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia."
DR EMBL: AF327054; AAK07469.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15234 MW; 03361A792F5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLTLT 170
DB 10 NLLATYGGPAGLTAFLRQIGDNTVTRLDRETELNEALPGDARTTTPASMAATLRLTLT 69
QY 171 GELLITLARSQQLIDMEADKVAAGPLLSALPAGWFIADKSGAGSGRGITIALGPDGKP 230
DB 70 SQRLSARSQRLQGMVDDRVAGPLIRSVLPAGWFIADKSGAGSGRGITIALGPNKKA 129
QY 231 SRIIVYI 237
DB 130 ERIIVLY 136

RESULT 11
Q9AMA3 PRELIMINARY; PRT; 139 AA.
ID Q9AMA3;
AC Q9AMA3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-935;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
RL Isolates from UMMC, Malaysia."
DR EMBL: AF327047; AAK07462.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.

```

```

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
Best Local Similarity 34.1%; Score 459; DB 2; Length 139;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGSPKELTAFLAHNGDHYTRLDMEPELNEAIPNDEDTTPVAMATTLKLLT 170
DB 10 HLLATVGGPAGLITAFNRICDNYTRLDMEPELNEAIPNDEDTTPVAMATTLKLLT 69
QY 171 GELLTASRQQLIDWMEADKVAQPLRLSALPAGWFIADKSGAGSGRSGLIATLGPDKP 230
DB 70 SORLSARSORQLQWVDDRVAGPLIRSVLPAGWFIADKSGAGSGRSGLIATLGPDKP 129
QY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 12
Q53553 PRELIMINARY; PRT; 67 AA.
AC 053553;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pHAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSW 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81098; AA835839.2;
DR HSSP: P00810; IXPB.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 67 AA; 7287 MW; 85F8C85B0363FCB CRC64;

Query Match
Best Local Similarity 25.1%; Score 338; DB 2; Length 67;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSRIDAQEQQLGRRIHYSQNDLVEYSPVTEKHLTDGMYRELCSAATMSQNTAANLL 114
DB 1 VLSRVDAQEQQLGRRIHYSQNDLVEYSPVTEKHLTDGMYRELCSAATMSQNTAANLL 60
QY 115 TTIGGPK 121
DB 61 TTIGGPK 67

RESULT 13
Q8GDE5 PRELIMINARY; PRT; 95 AA.
AC 08GDE5;

```

```

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Amyes S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
RT Chile."
RL Submitted (SEP-2002) to the EMBL/GenBank/DDB databases.
DR EMBL: AY150585; AAN7730.1;
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 24.2%; Score 326; DB 2; Length 95;
Matches 62; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 147 AIPNDEDTTPVAMATTLKLLTGELTLASRQQLIDWMEADKVAQPLRLSALPAGWFI 206
DB 1 ALPGARDTTPPASMATLRLKLTSSRFSORQLQWVDDRVAGPLIRSVLPAGWFI 60
QY 207 ADKSGAGSGRSGLIATLGPDKPSRIIVYITGGS 241
DB 61 ADKTGASERGARGIALLGPNKKAERTIVYIYLRDS 95

RESULT 14
Q9JNS8 PRELIMINARY; PRT; 62 AA.
AC 09JNS8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pHAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCSW 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL: S81099; AA835840.1;
DR HSSP: P00810; IXPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 23.0%; Score 310; DB 2; Length 62;
Matches 60; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 132 DHVTRLDWPEPLNEAIPNDEDTTPVAMATTLKLLTGELTLASRQQLIDWMEADKV 191
DB 1 DHVTRLDWPEPLNEAIPNDEDTTPVAMATTLKLLTGELTLASRQQLIDWMEADKV 60
QY 192 AG 193

```


THIS PAGE BLANK (USPTO)

SO SEQUENCE 176 AA; 18933 MW; ECE369602A0ABD02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 3e+02;

Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 IGGREKELTAFHNMGDHTRIDREPELENAIPNDERDTPVAMATTIRKLITGLT 176
DB 42 LGEKRVLAGFAGSVADAAATLDFRFAKLR-----WGNLTK 78
QY 177 ASROOLIDMWEADYAGPLRSALPAGFIADK-----SGAGEGSGNII-----AAL 224
DB 79 AAVELANDM-RTDRY-----LR-REALLVADKENIFLISNGE-----VIQPDDAAI 127
QY 225 GPDG 228
DB 128 GSGG 131

RESULT 14

ID PDUV_SALTY STANDARD; PRT; 150 AA.
AC Q9XDM6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Propanediol utilization protein pduV.
CN PDUV OR STM2056.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
NX NCBI_Taxid=602;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE-99429843; PubMed-10498708;
RA Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
RT "The propanediol utilization (pdu) operon of Salmonella enterica
serovar typhimurium lt2 includes genes necessary for formation of
polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-
propanediol degradation."
RL J. Bacteriol. 181:5967-5975(1999).
[2]

RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SCS1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;

RA McClelland M., Sanderson K.E., Spleth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
lt2."
RL Nature 413:852-856(2001).

CC -1- INDUCTION: By propanediol.
CC -1- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AF026270; MAD39020.1; -;
DR EMBL: AE008790; AAL20960.1; -;
DR StyGene: SG10680; pduV.
KM ATP-binding: Complete proteome.
FT NP_BIND 8 15 ATP (POTENTIAL).

SO SEQUENCE 150 AA; 16348 MW; 7771229432F97856 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 150;

Best Local Similarity 19.2%; Pred. No. 2.7e+02;
Matches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

QY 52 CGAVLSRIDAGQELGRIHYSONDLYEYSPYERKHLTDGWTVELCSAAITMSDNTAAN 111
DB 12 CG-----KTSLTQSIRGEALHKKFOAJEMSPMAID--TPGEYLENRCILYSALLTSACAD 65
QY 112 LLTTIGSPKELTAFHNMGDHTRIDREPELENAIPNDERDTPVAMATTIRKLITG 171
DB 66 VIALVNLNDQWSPF-----SGFTAPNMRPTIGLVTRADLAEPORISLVA 111
QY 172 ELUTIASROOLIDMWEADYAGPLRSALP 202
DB 112 EMLTQAGAQOLF-----ITSALNNGGLA 135

RESULT 15

ID YDEJ_ECOLI STANDARD; PRT; 172 AA.
AC P31131;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ydeJ.
GN YDEJ OR B1537.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NX NCBI_Taxid=562;
[1]

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-93186717; PubMed-8383113;
RA Cohen S.P., Haechele H., Levy S.B.;
RT "Genetic and functional analysis of the multiple antibiotic
resistance (mar) locus in Escherichia coli."
RL J. Bacteriol. 175:1484-1492(1993).
[2]

RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1234-1238(1997).
[3]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;
RX MEDLINE-97251357; PubMed-9097039;
RA Alba H., Bada T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).

CC -1- SIMILARITY: BELONGS TO THE CIMA FAMILY. STRONG, TO E. COLI YGAD.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M96235; -; NOT ANNOTATED_CDS.
DR EMBL: AE000251; AAC74610.1; -;
DR EMBL: D90796; BAA15227.1; -;

ID YCJC_ECOLI STANDARD: PRT: 185 AA.
 AC P38522; P76839; P77417;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein ycjC.
 GN YCJC OR B1299 OR SF1304.
 OS Escherichia coli, and
 OS Shigella flexneri.
 OC Bacteria: Proteobacteria: Gammaproteobacteria: Enterobacteriales;
 OC Enterobacteriaceae: Escherichia.
 NX NCBI_TaxId=562, 623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES-E.COLI; STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES-E.COLI; STRAIN-K12;
 RX MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashiwano K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sampei G., Seki Y., Sivasubram S., Yamagi H., Yareda J.,
 RA Takeuchi K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 65-185 FROM N.A.
 RC SPECIES-E.COLI;
 RX MEDLINE-91216440; PubMed-1840553;
 RA Helm R., Strehler E.E.;
 RT "Cloning an Escherichia coli gene encoding a protein remarkably
 similar to mammalian aldehyde dehydrogenases.";
 RL Gene 99:15-23(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES-S.flexneri; STRAIN-301 / Serotype 2a;
 RX MEDLINE-2272406; PubMed-12384590;
 RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: Insights into pathogenicity
 through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [5]
 RP IDENTIFICATION.
 RC SPECIES-E.COLI;
 RA Rudd K.E.;
 RL Unpublished observations (AUG-1994).
 CC -1- SIMILARITY: SOME, TO H.INFLUENZAE H10659.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000228; AAC74381.1; -
 CC EMBL: D90768; BAA14868.1; -
 CC EMBL: D90767; BAA14859.1; -

DR EMBL: M38433; -, NOT_ANNOTATED_CDS.
 DR EMBL: AE015157; AAN42915.1; -
 DR PIR: F64878; F64878.
 DR Ecogene: EG12431; ycjC.
 DR InterPro: IPR007113; Cupin_sup.
 DR InterPro: IPR001387; HTH_3.
 DR Pfam: PF01381; HTH_3; 1.
 DR SMART: SM00530; HTH_XRE; 1.
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 185 AA; 20091 MW; D7D0D3FD794A7768 CRC64;
 Query Match 4.7%; Score 63.5; DB 1; Length 185;
 Best Local Similarity 27.5%; Pred. No. 2.9e+02;
 Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;
 QY 145 NEAIPNDRDPTTPYAMATTLRLKLTGELTASRQQLIDWMEADKAGVPLRSALPAGM 204
 DB 34 HSAITIDQDKVSPA-ISTLQKLKLVGLSISE-----FSSEPKRPEPOV----- 78
 QY 205 FIADKSGAGERGSRGIIAALGPDGKPSRIYV-----YTTGSOATMDER-NRQIAETGASL 259
 DB 79 -VINODDILEMGSGVSMKLVHNGMNPRTIAMIFETYPGT--TTGERIKHGEIEIGVYL 135
 RESULT 13
 HSLV_THEME STANDARD: PRT: 176 AA.
 ID HSLV_THEME
 AC Q9WY21;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-dependent protease hslv (EC 3.4.25.-).
 GN HSLV OR TM0521.
 OS Thermotoga maritima.
 OC Bacteria: Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NX NCBI_TaxId=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 RX MEDLINE-99287316; PubMed-10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (by similarity).
 CC -1- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001728; AAD35606.1; -
 DR PIR: G72365; G72365.
 DR HSSP: P31059; 1E94.
 DR MEROPS: T01.006; -
 DR TIGR: TM0521; -
 DR HAMAP: MF_00248; -; 1.
 DR InterPro: IPR001353; Protease_protease.
 DR Pfam: PF00227; proteasome; 1.
 KW Hydrolase; Protease; Complete proteome.
 FT ACT_SITE 6 6 BY SIMILARITY.


```

OX NCBI_TaxID-2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C2A / ATCC 35395 / DSM 2834;
RX MEDLINE-21929760; PubMed-11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endlitz M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ancoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearcliano K., Johnson R.,
RA Lincon L., McEwan P., McKernan K., Talamas J., Tlrell A., Ye W.,
RA Zlmer A., Barber R.D., Cann I., Graham D.E., Grahme D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell R.F., Jing H., Macario A.J., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf M.W., Birren B.;
RT "The genome of Methanocaldococcus jannaschii reveals extensive metabolic
RT and physiological diversity.";
RI Genome Res. 12:532-542(2002).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
DR EMBL: AEC011164; AAM07774.1; -.
DR HAMAP: MF_00165; -.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR TIGRfam: TIGR00041; DTPM_kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; PALSE_NEG.
DR TRANSFAM: TRANSFAM_Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
KW FT_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C2E58C85423B3 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.7%; Pred. No. 2.3e+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

OY 56 ISRID-AGEOGLRRIHYSNDLYEVSPTKHLTDGMVRELCSAITSMTANILL 114
DB 8 LEGIDSGKSTVAKKIQ-ENSELRVEPVFTREPTKG-----TLTGAVERAIO 55
OY 115 TTIGPRELTAFLHNGDHYTRLDRWPELNEAIPNDRDRTTPVAMATTLKLTGELL 174
DB 56 SDTDQALAELEFLFADNAHEAKLVKPALEDEKTVISDSYAVAGTISLKNRLDNP 114
OY 175 TLASROQLIDMEADRYACPLRSALPACW-----FIAD-----KSG-AGE--- 214
DB 115 -----EMVR-----DLRGTWVLPDLTFLFDIPETIAVKKRGKRGQTKF 154
OY 215 -----RGSRGITIALGPOCKPSRIYVITGSOAIMDE 247
DB 155 EKIEPLRGVRELFLGLAAE-EPERFVAVDASGSPDEVER 192

RESULT 9
VNSC_RINDK STANDARD; PRT; 177 AA.
AC P35948;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein C.
GN C.
OS Rinderpest virus (strain Kabete O) (RDV).

```

```

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID-11242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92410646; PubMed-1529555;
RA Yamanaoka M., Dale B., Crisp T., Cordell B., Grubman M., Ylma T.;
RT "Sequence analysis and editing of the phosphoprotein (P) gene of
RT rinderpest virus.";
RI Virology 190:553-556(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
DR EMBL: S44819; AAB3269.1; -.
DR PIR: B43387; B43387.
DR InterPro: IPR003875; Paramyxovir_NSC.
DR Pfam: PF02725; Paramyxo_NSC; 1.
KW Nonstructural protein.
SQ SEQUENCE 177 AA; 19926 MW; 116971B140A39F11 CRC64;

Query Match 4.8%; Score 65; DB 1; Length 177;
Best Local Similarity 24.6%; Pred. No. 2.1e+02;
Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

OY 93 TVNELCSAAITNSDNTANILLTTIGPRELTAFLHNGDHYTRLDRWPELNEAIPNDE 152
DB 49 TIRISASHSQDLQAKACLAVTIRDEEATVAMS-----WEHSL----- 90
OY 153 ROTTTP-----VAMATTLKLTGELLTLASROQLIDMEADRYACPLRSALPA 202
DB 91 ----VTTCOAPRYSTIMFTAVKRLRESKMLTSLWFNOL--MMVSK-SGEBMRRLRTA 144
OY 203 GWFID 208
DB 145 MWLIAN 150

RESULT 10
VNSC_RINDR STANDARD; PRT; 177 AA.
AC 003339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Nonstructural protein C.
GN C.
OS Rinderpest virus (strain RBOK) (RDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID-36409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93155658; PubMed-8429304;
RA Barton M.D., Shalla M.S., Barrett T.;
RT "Cloning and sequence analysis of the phosphoprotein gene of
RT rinderpest virus.";
RI J. Gen. Virol. 74:299-304(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
DR EMBL: X68311; CAA48391.1; -.

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR U0321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=2050219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caspell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AE002129; AAF30730.1;
DR HAMAP: MF_00003; -.
DR InterPro: IPR000238; Rib_bind_facta.
DR Pfam: PF02033; RBFA; 1.
DR ProDom: PD007327; Rib_bind_facta; 1.
DR TIGRFAMs: TIGR00082; rbfA; 1.
DR PROSITE: PS01319; RBFA; FALSE NEG.
KW rRNA processing; Complete proteome.
SQ SEQUENCE 116 AA; 13247 MW; A79DC71F0547514 CRC64;

Query Match 4.9%; Score 66.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. NO. 91;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

OY 144 LNEALPDERPTTPVAMATTLKILGEL-----ITLASRQQLIDME-ADRYAGPL 196
DB 18 INMLANEINDKIARLAVR--LSNDLSVAKIFDAHRRESKLVLENVKNVSG-L 74
OY 197 RSALPAGW-----PIADRS 210
DB 75 RSKLAEMTSYKVPFLRFVIDET 97

RESULT 7
OGG1_ARCFU STANDARD: PRT; 198 AA.
ID O29876;
AC 029876;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable N-glycosylase/DNA lyase [Includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN OGG OR APO371.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

```

```

RX MEDLINE=98049343; PubMed=9389475;
RA Klek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwyn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Keriavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
CC guanine (7,8-dihydro-8-oxoguanine - 7-oxoG) from DNA. Also nicks
CC DNA at apurinic/aprimidinic sites (AP sites) (By similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL: AE001079; AAB90876.1;
DR PIR: C69296; C69296.
DR TIGR: AF0371; -.
DR HAMAP: MF_00241; -.
DR InterPro: IPR003265; Endo_3c.
DR Pfam: PF00730; HNH-GPD; 1.
DR SMART: SM00478; ENDO3C; 1.
DR HydroLase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
KW Multifunctional enzyme; Complete proteome.
FT ACT SITE 122
FT SEQUENCE 198 AA; 22639 MW; 3A5C03AA12F3FEB CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 198;
Best Local Similarity 25.7%; Pred. NO. 2.2e+02;
Matches 37; Conservative 19; Mismatches 53; Indels 35; Gaps 8;

OY 14 QLGARVGTIEDLSGKILSFREPERPMSTFKVLGAVL--SRIDAG---QEOLG 67
DB 15 QLGER-GEVERD-----FRPFLDFSVKATIRTELAFCSIPANSATAGLKFQRLG 64
OY 68 RRIHYSQ-----NDLVEYSPYTEKHLTDGMYRELCSALITSDNTAALLTTI 117
DB 65 QGVGKELTLTAGYRFHNRKAEY-----IREAFKSFILVEKALESSKAREILTKIK 117
OY 118 G-GPKETLAFLHNG-DHYTRIDR 139
DB 118 GIGMKKASHFLRNVGREDVALIIDR 141

RESULT 8
KTHY_METAC STANDARD: PRT; 206 AA.
ID Q87HS9;
AC Q87HS9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR MA4433.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.

```

QY 100 AATMSDNTAANLLTTIGPKK-----LTAF-LHNMGDHYRLDRWEDE 143
 DB 45 ANVYV-DGSTVNLGLMDTAGOEDYNNRLPLSYRGADVFLFASLSRASENISKWIE 103
 QY 144 LNEAINDERDOTTTPAAMATTLLKLLTGELLLT-ASRQOLIDMEADKVA---GPIILRSA 199
 DB 104 LRHYAV-----TVPI-----VLVGTRLDREDROYLIDHPGATPTTAGGEELKRA 149
 QY 200 LPAGWFI 206
 DB 150 IGAAYVL 156

RESULT 4

NDK_TREPA STANDARD: PRT: 149 AA.
 ID NDK_TREPA
 AC 083974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 GN NDK OR TP1010.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 NC NCBL_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Welnslock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardam J.M., McLeod M.P., Salberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
 RA McDonald L., Attleach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 agent Spirochaete";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate -> ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL: AE001269; AAC65961.1; -
 DR PIR: F71252; F71252.
 DR HSSP: P22887; INPK.
 DR TIGR: TP1010; -
 DR HAMAP: MF_00451; - 1.
 DR InterPro: IPR001564; NDK.
 DR Pfam: PF00334; NDK.1.
 DR PRINTS: PR01243; NUCDPKINASE.
 DR PRODOM: PD001018; NDK.1.
 DR SMART: SM00562; NDK.1.
 DR PROSITE: PS00469; NDP_KINASES; 1.
 KW Transferrase; Kinase; ATP-binding; Complete proteome.
 FT ACT_SITE 117 117 By similarity.
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

Query Match 5.0%; Score 68; DB 1; Length 149;
 Best Local Similarity 21.6%; Pred. No. 95;

Matches 35; Conservative 25; Mismatches 64; Indels 38; Gaps 7;
 QY 29 GKILSEFRPEPRPMSTFKVLICGAVLSRIDAGOBOLRRIRYSONDLVESPTEKHL 88
 DB 21 GEVLSRF---ERKGLVLTALRLC-----VDATREL---HYAHRKRPFLSLIAYI 67
 QY 89 TDGMTVRELCSAATMSDNTAANLLTTIGPRELTAFLHMGDHYTRLDRWPEPELNEAI 148
 DB 68 TSAFVY-----ALAFKEMALISVRLCS-----TVREHAQP---GFI 103
 QY 149 PND-ERDITTPVAMATTLLKLLTGELLLTASRQOLIDMEAD 189
 DB 104 RGDFALTRTTNTIVHSDSPESARELALYFSADQFVEMRGN 145

RESULT 5

HC3L_THIFE STANDARD: PRT: 122 AA.
 ID HC3L_THIFE
 AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (hydrogenase)
 GN (fragments).
 DE HOXG.
 OS Thiodacillus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 NC NCBL_TaxID=920;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=ATCC 19859;
 RX MEDLINE=96241863; PubMed=8661919;
 RA Fischer J., Quantenler A., Kostka S., Kraft R., Friedrich C.G.;
 RT "Purification and characterization of the hydrogenase from
 RT Thiodacillus ferrooxidans";
 RL Arch. Microbiol. 165:289-296(1996).
 CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 + 2
 CC ferrocyclochrome c3.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 CC
 KW Oxidoreductase; Iron.
 FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 78 79
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EB4FAF365348DA0B CRC64;

Query Match 5.0%; Score 67; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 89;
 Matches 28; Conservative 13; Mismatches 23; Indels 42; Gaps 4;

QY 19 VGTELDLNSGKILSEFRPEPRPMSTFKVLICGAVLSRIDAGOBOLRRIRYSONDLV 78
 DB 12 VGRVEGLDLSXSLIEYFR-----NATLAFRGG-----LC 41
 QY 79 EYSPVTEKHLTDGMTVRELCSAATMSDNTAANLLTTIGPRELT 124
 DB 42 KTAFTGTNTEIGVT-----ISGDKDQAGLIVT-----PREST 75

RESULT 6
 RBFA_UREPA STANDARD: PRT: 116 AA.
 AC Q9P0H0;

KW Hypothetical protein.
SQ SEQUENCE 191 AA; 21549 MW; 65555347F0EBED16 CRC64;

Query Match 5.7%; Score 77; DB 1; Length 191;
Best Local Similarity 22.6%; Pred. No. 24;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;

QY 104 MSDNTAANLLTTIGPKELTAFHNMGDHTRLDREWEDELMKAIIPNDERDTTPVAMAT 163
DB 1 MSYMSLNLMLONPSGIDKIAAILVN---VARLD-----PASKSTPAQLVSMLN 46
QY 164 TLRLKLTGELTLASRQOLIDWMEADKYAGPLLRASALPAGPIADKSGGEGSGGIIAA 223
DB 47 EERC-----ILRLPGLYKILIVFRDSSPEYTMNAINIGYVTE--GLAIFGKGQIIISI 99
QY 224 LGPDGKP-----SRIVIYTTGSGQATPDERNRQI----- 252
DB 100 ----SKPLEDKLMLWSRFFMLDPLTLYQLLRKTEDEKREKQDLASNLASLPICIMWS 155
QY 253 AEIGASLTKH 262
DB 156 VENGAGLKH 165

RESULT 2

RAC2_DICDI
ID RAC2_DICDI STANDARD; PRT; 201 AA.
AC 09G9S0; 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RAS-related protein rac2.
GN RAC2.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Eukaryota; Dictyosteliales; Dictyostelium.
OX NCBI_Taxid=44689;
RN NCB1
RP SEQUENCE FROM N.A.
RC STRAIN-AX4;
RX MEDLINE=21127961; PubMed=11222756;
RA Rivero F., Dislich H., Glockner G., Noegel A.A.;
RL Nucleic Acids Res. 29:1068-1079(2001).
CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF310893; AAG45130.1; -
DR HSSP; P21181; 1M44.
DR DictyDB; DD77777; racg.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR001806; Ras_trnsfmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00174; RHO; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 57 61 GTP (BY SIMILARITY).
FT NP_BIND 115 118 GTP (BY SIMILARITY).
FT DOMAIN 32 40 GTP (BY SIMILARITY).
FT LIPID 198 198 GERANYL-GERANYL (POTENTIAL).
SQ SEQUENCE 201 AA; 22813 MW; F6CE586497CC169 CRC64;

Query Match 5.6%; Score 75; DB 1; Length 201;
Best Local Similarity 22.6%; Pred. No. 38;

Matches 36; Conservative 20; Mismatches 59; Indels 44; Gaps 7;

QY 44 MSTFVLLCGAVLSRIDAGQDGLRRIRHSQNDLV-EYSP-VTEKHLDGTVRELCSAA 101
DB 1 MSIVVCYVG-----EGIGKTSMLLSYTSNISSENPQVFDNYST----- 42
QY 102 ITMSDNTAANLLTTIGPKELTAF-----LHNMGDHTRLDREWEDELM 145
DB 43 LHMNRKRYNLSINDTQAQEEFSKRLRLSTPQTDVFLCFLINDPSFSNILDSDNVDRLN 102
QY 146 FAIPNDERDTTPVAMATTLKRLTGLLEL--TLASRQOL 182
DB 103 ENCPN-----TPILVCTQMDLKSNSVILRLCEKQKL 135

RESULT 3

RAC2_LOTJA
ID RAC2_LOTJA STANDARD; PRT; 196 AA.
AC 040220; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAC2-like GTP binding protein RAC2.
GN RAC2.
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
OX NCBI_Taxid=34305;
RN NCB1
RP SEQUENCE FROM N.A.
RC STRAIN-cv. Gifu / B-129; TISSUE=Root nodules;
RX MEDLINE=97231679; PubMed=9076991;
RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
RL Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from Lotus japonicus, and expression of corresponding mRNAs in developing root nodules.*;
RL Plant J. 11:237-250(1997).
CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z73962; CA98190.1; -
DR HSSP; P21181; 1M44.
DR InterPro; IPR003578; GTPase_Rho.
DR InterPro; IPR001806; Ras_trnsfmg.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00071; ras; 1.
DR PRINTS; PR00449; RASTRNSFRMG.
DR SMART; SM00174; RHO; 1.
DR TIGRfams; TIGR00231; small_gtp; 1.
KW GTP-binding; Prenylation; Lipoprotein.
FT NP_BIND 13 20 GTP (BY SIMILARITY).
FT NP_BIND 60 64 GTP (BY SIMILARITY).
FT NP_BIND 118 121 GTP (BY SIMILARITY).
FT DOMAIN 35 43 GTP (BY SIMILARITY).
FT LIPID 193 193 GERANYL-GERANYL (POTENTIAL).
SQ SEQUENCE 196 AA; 21704 MW; 8BB0A70C4828CD5 CRC64;

Query Match 5.1%; Score 68.5; DB 1; Length 196;
Best Local Similarity 24.1%; Pred. No. 1.2e+02;
Matches 45; Conservative 23; Mismatches 64; Indels 55; Gaps 11;

QY 44 MSTFVLLCGAVLSRIDAGQDGLRR--IHSQNDLV-EYSPVTEKHLDGTVRELCS 99
DB 1 MSTAFIKC-----VTVDGAGVGTGCMILSYTSNFPFDYV-----TVFDNFS 44

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 5.57143 seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_157T

Perfect score: 1348

Sequence: 1 HPELVKVKDAEDQGARVG.....TMDERNROIAEIGASLIKHM 263

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt-41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	77	5.7	191	YDA4_SCHPO	Q10346 schizosacch
2	75	5.6	201	RACG_DICDI	Q99800 dictyostell
3	68.5	5.1	196	RAC2_LOTJA	Q40220 lotus japon
4	68	5.0	149	NDK_TREPA	Q83974 treponema p
5	67	5.0	122	HC3L_THIFE	P80509 thiodicilla
6	66.5	4.9	116	HBFA_UREFA	Q99810 ureaplasma
7	65.5	4.9	198	OCGI_ARCFU	Q29876 archaeoglob
8	65.5	4.9	206	KTHI_METAC	Q81859 metanostarc
9	65	4.8	177	VNSC_RINDR	P35948 rinderpest
10	65	4.8	177	VNSC_RINDR	Q03349 rinderpest
11	63.5	4.7	144	MARR_ECOLI	P27245 escherichia
12	63.5	4.7	185	YCJC_ECOLI	P38522 escherichia
13	63	4.7	176	HSIV_THEMA	Q99871 thermotoga
14	62.5	4.6	150	PDUV_SALT	Q9X466 salmonella
15	62.5	4.6	172	YDEB_ECOLI	P31131 escherichia
16	62.5	4.6	182	PYRE_STRCO	Q9X817 streptomyce
17	62.5	4.6	195	TRPF_THEVO	Q97946 thermoplasm
18	62.5	4.6	200	RR4_PELNE	Q9F499 pellicula nees
19	62	4.6	174	THBB_RAT	P17481 rattus norv
20	62	4.6	182	YH61_BIFLO	Q89520 bifidobacte
21	62	4.6	186	YCEB_SALT	P40822 salmonella
22	61.5	4.6	178	HSIV_RALSO	Q8Y347 ralsomonia s
23	61	4.5	121	SECR_HUMAN	P09683 homo sapien
24	61	4.5	200	TATR_CANCR	Q9A611 caulobacter
25	61	4.5	202	COAT_ELY	P35927 crysallum la
26	60.5	4.5	103	RS10_NERGO	P48851 neisseria g
27	60.5	4.5	146	HBG_RABIT	P02099 oryctolagus
28	60.5	4.5	156	BFR_AZOVI	P22759 azotobacter
29	60.5	4.5	160	TATB_HELPY	Q25700 helicobacte
30	60.5	4.5	184	MLRI_SCHPO	Q9U055 schizosacch
31	60.5	4.5	193	SAR2_LYCES	P52884 lycopersico
32	60.5	4.5	196	ALKH_BACSU	P50846 b.kh/kdpg
33	60	4.5	148	CALM_BLAEM	Q9H1Y6 blaetoclad

34	60	4.5	159	1	GREX_CHLTE	Q8KX5 chlorobium
35	60	4.5	161	1	PIN_BPT4	P07068 bacterioph
36	60	4.5	178	1	PYRE_ARCFU	Q28533 archaeoglob
37	60	4.5	178	1	UCRI_ANASP	P70758 anabaena sp
38	60	4.5	184	1	ARL2_DROME	Q06849 dirosophila
39	60	4.5	205	1	RS2_ASPEP	Q9YB45 aeropyrum p
40	59.5	4.4	103	1	RS10_NEIMA	Q9J121 neisseria m
41	59.5	4.4	135	1	CCRN_PAROL	Q57312 paracitichth
42	59.5	4.4	184	1	HRPL_PSESY	P37929 pseudomonas
43	59.5	4.4	196	1	I196_ASFB7	P27943 african swi
44	59.5	4.4	198	1	UPP_ARATH	Q65583 arabidopsis
45	59.5	4.4	205	1	RS4_ECOLI	P02354 escherichia

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD; PRT; 191 AA.
ID YDA4_SCHPO
AC Q10346:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F12.04c in chromosome 1.
GN SPAC1F12.04c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks R., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsals K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grignonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabell C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Beger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaur V., Mottier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Koehn M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe".
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL: Z69944; CAA93808.1; -
CC PIR: T38062; S67447.
CC GeneDB_Spombe; SPAC1F12.04c; -

THIS PAGE BLANK (USPTO)

A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0825
C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match
Best Local Similarity 5.3%; Score 71; DB 2; Length 177;
Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

DB 42 PMSTFKVLLCGAVSRIDAGEQGRRIHYSQNDLVEYSPTEKHLTDGMTVRELCSA 101
DB 11 PEMDTSVGLFYAMVEENVND-----RLH-----HLIDVTEELLYKG 47
DB 102 ITMSDNTAANLLTTTGPKELTAFLHNMGDVTRL-DWEPEL-NEAIPND----- 151
DB 48 SDGDESMAGLL-----NHLYTVDAWVPRINGEALPDSLEAHGPM 89
DB 152 -ERDTPPYAMATTLTKLITGELTLASNO-----LIDMEADKVACPLRLSALP 201
DB 90 VKDKGLPVYVLSVDELIEKORYVALIKETCALHDDDLARWIPYEEROATIRWGL- 148
DB 202 AGMFTAD 208
DB 149 --WHMAD 153

RESULT 11

conserved hypothetical protein PA4399 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
C:Accession: G83096

R:Stover, C.K.; Pham, X.Q.; Eryn, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bredman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardi, K.; Lam, N.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: G83096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <STO>
A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA07787.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4399
C:Superfamily: conserved hypothetical protein DR1638

Query Match
Best Local Similarity 5.3%; Score 71; DB 2; Length 192;
Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

DB 53 GAVLSRI-----DAGQEDL--GRIHYSQNDLVEYSPTEKHLTDGMTVRELCSA----- 100
DB 2 GNRLSKYTRTGDSGEGTLAGRRVPSKPRITGAIDVDELNSQGLLALLLEARGAP 61
DB 101 -----AITSMDTAAANLLTTTGPKELTAFLHNMGDVTRL-----DWEPEL-----N 145
DB 62 GLEIVQALAPVCR-----LFDJGGLAMEYRALDETETVARLESIDRNDELCPKLN 116
DB 146 EAIPIIDERDTTPVAMATTLTKLITGELTLASRQDLIDMEADKVAGPLRLSALPAGWF 205
DB 117 FILGGSR-----PVAQAHVCRSLAR-----SAERRQALDDEETIEGVLRIANRLSDLLF 168
DB 206 IADKSGAGERSRRII--AALGPD 227
DB 169 VAARAIARQGVAAEILWEAAKPD 192

RESULT 12

conserved hypothetical protein Yng0594h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84217
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
Jung, K.H.; Alam, M.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.B.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483; PMID:11016950
A:Accession: E84217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <STO>
A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA019105.1; GSPDB:GN0138
C:Genetics:
A:Gene: Yng0594H

Query Match
Best Local Similarity 5.2%; Score 70; DB 2; Length 108;
Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

DB 44 MSTFVLLCGAVSRIDAGEQGRRIHYSQNDLVEYSPTEKHLTDGMTVRELCSA 103
DB 1 MDTVTELDGALIALEATD-----RVEFRFDALVTDVTLRFHDDDRV-----GSIT 50
DB 104 MSDNTAANLLTTTGPKEL-----TAFIHMGDVTRLDR 139
DB 51 NDDGTRTMAKLTVPDSDFLAVEPTSFVAIVDAARTRDR 92

RESULT 13

conserved hypothetical protein YP01575 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AB0192

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-170 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90397.1; PID:g15979615; GSPDB:GN0175
C:Genetics:
A:Gene: YP01575
C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match
Best Local Similarity 5.2%; Score 70; DB 2; Length 170;
Matches 22; Conservative 25; Mismatches 44; Indels 14; Gaps 3;

DB 56 LSRIDAGEQGRRIHYSQNDLVEYSPTEKHLTDGMTVRELCSAATMSDNTAANLLT 115
DB 70 MNKIDSKGNIG--FNFTDNL-----KTIIVDKLTAQLIGSLAIA-----RLVV 115
DB 116 TIGPKELTAFLHNMGDVTRLDRWEPELNEAIPNDERDTTPVA 160
DB 116 DNGSESEVAITPASVADKIAORDASSIVLSALSIOBQDDEDPYA 160

RESULT 14

conserved hypothetical protein Yng0953c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84251
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
Jung, K.H.; Alam, M.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja

Best Local Similarity 22.6%; Pred. No. 34;
Matches 43; Conservative 22; Mismatches 69; Gaps 7

[illegible]

RESULT 3

hypothetical protein u1756f [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robinson, K.
Submitted to the EMBL Data Library, September 1994
A:Reference number: Z16911
A:Accession: T45195
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:015180, PIDN:AAA62885.1

```

Query Match      5.58; Score 74.5; DB 2; Length 113;
Best Local Similarity 28.68; Pred. No. 27;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;

QY 130 MGDHV--TRLDREPELN--EAIPNERDT---TPVMAATLRKLINGELLTLASRQ 180
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1 MGGGIGNERGSRNTGNTQCPRLRVPGESPTLDGRASPEDIIT--NLTSTPMISHPPS 58

QY 181 QLIDMKEA-DKVGAPLL-----NSALPAGMTLADKSGAGERSGIIAALPGGKPSR 232
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 59 RDDDWVEPFDALGTAVFATGKATMPANVGGI---GASTRGS-GILASISPPROPAR 112

```

RESULT 4

molybdopter in biosynthesis mog protein [Imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_rev:10n 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AD3606
 R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivancov, R.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leclerc, N.A.L. Acad. Sci. U.S.A. 99, 443-448, 2002
 A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
 A:Reference number: AD3352; PMID:11755688
 A:Accession: AD3606
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1172 <RUR>
 A:Cross-references: GB:EM08918; PIDN:AM54015.1; PID:917984966; GSPDB:GM00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI10773
 A:Map position: II
 C:Superfamily: molybdenum cofactor biosynthesis protein B moaB

Query Match	5.58;	Score 74.5;	DB 2;	Length 172;
Best Local Similarity	25.58;	Pred. No. 48;		
Matches	40;	Conservative 14;	Mismatches 42;	Indels 61;
				Gaps 9;

[illegible]

RESULT 5

hypothetical protein SCJ11.13 - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C/Accession: T365984
R/Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z21618
A/Accession: T365984
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-152 <OL1>
A/Cross-references: EMBL:AL109949; PIDD:CA852898.1; GSPDB:GN00070; SCOEDB:SCJ11.13
A/Experimental source: strain A3(2)
C/Genetics:
A/Gene: SCOEDB:SCJ11.13
C/Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

[illegible]

RESULT 6

ADP-ribosylation factor homolog ARL3 - Trypanosoma brucei
N/Alternate names: ADP ribosylation factor 3 homolog
C/Species: Trypanosoma brucei
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C/Accession: T09136
R:Bruggaud, F.; Vedetne, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, M.O. Biochem. Parasitol. 94, 249-264, 1998
A/Title: Conserved organization of genes in trypanosomatids.
A/Reference number: Z16580; MVID:98418771; PMID:9747975
A/Accession: T09136
A/Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-190 <BRI>
A/Cross-references: EMBL:AF031926; NID:93452211; PIDN:MAC32774.1; PID:93452215
A/Experimental source: strain AntAt1
A/Note: small G-protein
C/genetics:
A:Gene: ARL3
C:Superfamily: ADP-ribosylation factor
C/Keywords: blocked amino end, lipoprotein, myristylation; nucleotide binding; P-loop
P124-31/region: nucleotide-binding motif A (P-loop)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_157T
Perfect score: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDERNROIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524	38.9	105	2	JC2566 bla protein - Pseu
2	77	5.7	191	2	S67447 hypothetical prote
3	74.5	5.5	113	2	T45195 hypothetical prote
4	74.5	5.5	172	2	AD3606 molybdopterin bios
5	71.5	5.3	152	2	T36984 hypothetical prote
6	71.5	5.3	190	2	T09136 ADP-ribosylation f
7	71.5	5.3	195	2	T36975 hypothetical prote
8	71.5	5.3	200	2	G97064 spore coat protein
9	71	5.3	153	2	E81708 conserved hypotet
10	71	5.3	177	2	A83753 hypothetical prote
11	71	5.3	192	2	G83096 conserved hypotet
12	70	5.2	108	2	E84217 conserved hypotet
13	70	5.2	170	2	AB0192 conserved hypotet
14	68.5	5.1	145	2	FB4251 conserved hypotet
15	68.5	5.1	167	2	D87360 conserved hypotet
16	68.5	5.1	192	2	A83587 hypothetical prote
17	68	5.0	149	2	F71252 nucleoside-diphosp
18	67.5	5.0	131	2	AD2281 hypothetical prote
19	67.5	5.0	184	2	T21126 ADP-ribosylation f
20	66.5	4.9	116	2	C82906 ribosome binding f
21	66.5	4.9	177	2	D90227 conserved hypotet
22	66.5	4.9	198	2	D95285 conserved hypotet
23	65.5	4.9	42	2	F56978 collagen alpha 1(X
24	65.5	4.9	181	2	AB1902 hydrolase small
25	65.5	4.9	195	2	AE0623 probable bacteriop
26	65.5	4.9	198	2	C69296 conserved hypotet
27	65	4.8	148	2	E75283 conserved hypotet
28	65	4.8	150	2	T08585 calmodulin - soybe
29	65	4.8	160	1	E69186 conserved hypotet

30	65	4.8	168	2	B75498 conserved hypotet
31	65	4.8	177	1	B43387 nonstructural prot
32	65	4.8	177	2	J01931 nonstructural prot
33	65	4.8	180	2	C71869 hypothetical prote
34	65	4.8	180	2	G70912 hypothetical prote
35	64.5	4.8	145	2	H75262 hypothetical prote
36	64.5	4.8	151	2	D81333 probable protein-t
37	64	4.7	113	2	D70580 hypothetical prote
38	64	4.7	162	2	AG0769 probable acetyltra
39	64	4.7	178	2	I40124 outer surface prot
40	64	4.7	191	2	E95333 hypothetical prote
41	63.5	4.7	164	2	H82336 regulator of sigma
42	63.5	4.7	166	2	C90029 hypothetical prote
43	63.5	4.7	168	2	T20606 hypothetical prote
44	63.5	4.7	179	2	AB1994 hypothetical prote
45	63.5	4.7	185	2	D90863 hypothetical prote

ALIGNMENTS

RESULT 1

JC2566
bla protein - Pseudomonas aeruginosa (fragment)
C:Species: Pseudomonas aeruginosa
C:Date: 26-Jun-1995 #sequence_rev1510 14-Jul-1995 #text_change 03-May-1996
C:Accession: JC2566
R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 14811281, 81-86, 1994
A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A:Note: due to a typographical error the volume number 148 appears as 128
A:Accession: JC2566
A:Molecule type: DNA
A:Residues: 1-105 <WES>
C:Genetics:
A:Gene: bla
C:Superfamily: beta-lactamase I

Query Match 38.9%; Score 524; DB 2; Length 105;
Best local Similarity 100.0%; Pred. No. 2.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMTTRKLLTGELTLASROQLIDMMEADKVAAGPLRBNALPAGWPIADKSGNGEGRSG 219
DB 2 AAMTTRKLLTGELTLASROQLIDMMEADKVAAGPLRBNALPAGWPIADKSGNGEGRSG 61
QY 220 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 263
DB 62 IIALGPDGKPSRIIVYITTSQATMDERNROIAEIGASLIKHW 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 20-Jul-1996 #sequence_rev1510 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: T38062; S67447
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, March 1996
A:Reference number: Z21766
A:Accession: T38062
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <MC2>
A:Cross-references: EMBL:Z69944; NID:g1217974; PIDN:CAA93808.1; PID:g1217978; GSPDB:G
C:Experimental source: strain 972n-; cosmid c1F12
C:Genetics:
A:Gene: SPAC1F12.04c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c
Query Match 5.7%; Score 77; DB 2; Length 191;

```

      RESULT 15
      US-10-156-761-12656
      ; Sequence 12656, Application US/10156761
      ; Publication No. US20030119018A1
      GENERAL INFORMATION:
      APPLICANT: OMDRA, SATOSHI
      APPLICANT: IKEDA, HARUO
      APPLICANT: ISHIKAWA, JUN
      APPLICANT: HORIKAWA, HIROSHI
      APPLICANT: SHIBA, TADAYOSHI
      APPLICANT: SAKAKI, YOSHIYUKI
      APPLICANT: HATTORI, MASAHIRA
      TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
      FILE REFERENCE: 249-262
      CURRENT APPLICATION NUMBER: US/10/156,761
      CURRENT FILING DATE: 2002-05-29
      PRIOR APPLICATION NUMBER: JP 2001-204089
      PRIOR FILING DATE: 2001-05-30
      PRIOR APPLICATION NUMBER: JP 2001-272697
      NUMBER OF SEQ ID NOS: 15109
      SEQ ID NO 12656
      LENGTH: 195
      TYPE: PRT
      ; ORGANISM: Streptomyces avermitilis
      US-10-156-761-12656

      Query Match          4.8%; Score 64.5; DB 15; Length 195;
      Best Local Similarity 22.1%; Pred. No. 1.4e+02;
      Matches 50; Conservative 21; Mismatches 88; Indels 67; Gaps 9

      QY      16 GARGVYIIDLNSGRI-----LESFPERPERPMWSTFKVLGCA-----VLSRIDAG-- 62
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      QY      4 GTAAVASYLRLDPSGICLDLATTTPRERLDSADLRAMATGSGVLPCTPPLPADASWP 63
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      QY      63 -----QQLGRRIHYSQNDLYEVSPTVEKHILTDGMTVRELCSSAITSNDTNTANILLTT 116
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      DB      64 VRFELRLQIGQLVHGE---LEGRPV-----DALATRYNVA 97
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      QY      117 IGGREFLTAFLLNNMGDHTVRLDRWEPELENEAIPNDRDQTTTPVAMATTLRLKLLGELLTL 176
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      DB      98 LGSPPICAVRADDGVLVRLPDA-APGCGELLAAIARTV-----ELLTD 141
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      QY      177 ASROOLIDMMADKAVAGFLLSALPAG---WFIADKSGAGERGRS 218
      DB      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
      DB      142 PAARASLRHCEDNMC--PLVYLDTSRGRRRRCSSSETGCNEERVAR 185

```

Search completed: September 10, 2003, 12:33:18
Job time : 17.5714 secs

Db 61 GELFMAEORGELNFGAGYTLFVSPGVNDPL--TLPRHW-----RRNGRLVA 107
 QY 224 LGPDGKP--SRIVVYTTGSGATMDERNROIAIGASLUKH 262
 Db 108 LKQDQKKRYARIGALMDAETHVLIERNKKLLEVTTLLEAH 148

RESULT 12

US-10-156-761-8623
 ; Sequence 8623, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 8623
 ; LENGTH: 194
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-8623

Query Match 5.0%; Score 67; DB 15; Length 194;
 Best Local Similarity 38.6%; Pred. No. 79;
 Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

QY 202 AGWFIADKSGAGRGSGRG---TIALGPDG--KPSRIYVYVT 239

Db 75 SGRVWTPRSAGSGRVTGAGYRTTAIGPLAVHEPVRAVVVVT 118

US-10-127-816-11
 ; Sequence 11, Application US/10127816
 ; Publication No. US20030104416A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Kluecher, Kevin M.
 ; APPLICANT: Taft, David W.
 ; APPLICANT: Kindsvogel, Wayne R.
 ; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
 ; FILE REFERENCE: 01-17
 ; CURRENT APPLICATION NUMBER: US/10/127,816
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/285,408
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: US 60/286,482
 ; PRIOR FILING DATE: 2001-04-25
 ; PRIOR APPLICATION NUMBER: US 60/341,050
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 60/341,105
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 09/895,834
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/285,424
 ; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 202

; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-127-816-11

Query Match 4.9%; Score 66; DB 15; Length 202;
 Best Local Similarity 21.5%; Pred. No. 1,4e+02;
 Matches 41; Conservative 23; Mismatches 65; Indels 62; Gaps 8;

QY 50 LILGAVLSRIDAGOEOLGRIRHYSOND--LVEYSPYTERKHL-----TDGMFVREL----- 97
 Db 17 LLLAAVLTRTQADPPVPRATRLPVEAKDCHIAQFSLSPRELQAFKAKGAIERKLENDM 76
 QY 98 -CSA-----ATMS--DNTAALLTTIGGPRELAFELN 129
 Db 77 RCSSHLISRAWDLKOIQVGERPKALQAEVALTLKWEININDSALTITIGQPLHTLSIHS 136
 QY 130 MGDHVT-----RLDRWEPELNEA---IPNDERDTTPVAMATTLRLKLTGE 172
 Db 137 QIQTCTQLOATYAPPPRPSRLSRWLRLQEAOSKETPGCLEDSVT-----SNLFQULLRD 191
 QY 173 LUTLASROOLI 183
 Db 192 LKCVASGDQCV 202

RESULT 14

US-10-156-761-8978
 ; Sequence 8978, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 8978
 ; LENGTH: 191
 ; TYPE: PRT
 ; ORGANISM: Streptomyces avermitilis
 ; US-10-156-761-8978

Query Match 4.8%; Score 64.5; DB 15; Length 191;
 Best Local Similarity 21.9%; Pred. No. 1.4e+02;
 Matches 44; Conservative 24; Mismatches 62; Indels 71; Gaps 10;

QY 54 AVLSRIDAGOEOLGRIRHY--SONDLVEYSPYTEKHLTDGMVRELCSAATITSDTAN 111
 Db 32 AILERAGAGQ---GSMVHREKGPDL-----ALTAIRRTAE 65
 QY 112 LLTTTGGPRELAFELNMGDHTRLDRWEPELNEAIPNDERDTT--TPVAMATTLRLK 169
 Db 66 LRAAESA-----LSGPTPYERI-----EAVLRERALKGCPIGRLTMDPDV 110
 QY 170 TGEILLTASROOLDW-----MEADKVGCPULRSALPAGFTAD 208
 Db 111 ASEELR-APVDEITIDMLREGLAIVEGKRGQFAGSLDAEELAAITL-ATVGGGYLAR 168
 QY 209 KSG---AGERGSGITIAALCP 226
 Db 169 AAGSPAFAFDYGVGLLSLRP 189

```

Query Match      5 0%: Score 67; DB 10; Length 178;
Best Local Similarity 24.8%: Pred. No. 70;
Matches 25; Conservative 17; Mismatches 37; Indels 22; Gaps 4

0Y      171 GELLTFLASROOLDIW-----MEADYVAGLLLSALPAGWFLADNRSGAGEGRSRITIAA 223
      ||| | : : : : : | | | | : || : |

```


OY 135 TRIDRWEPELNEAIPNDERDTTPVAMATT 164
DB 61 TRIDRWEPELNEAIPNDERXTIPVAMATT 90

RESULT 2

US-10-127-816-9
; Sequence 9, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Kluehner, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-9

Query Match
Best Local Similarity 5.6%; Score 75; DB 15; Length 202;
Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LILGAVLSRIDAGQGLRRIRHYSOND--LVEYSPTTEKHL-----TDGMTVREL----- 97
DB 17 LLLAAVLTQTQADPVPRAIRLPEVKAQCHIAQFKSLSPRELOAFKKAQDAIEKRLERDL 76
OY 98 -CSA-----AITMS--DNTAANLLTTTIGGPKELTAFLAN 129
DB 77 RCSSHLFPRAWDKQLOQVERPKALQAEVALTLKWMWTDALATITIGQPLHTLSHTS 136
OY 130 MGDHYT-----RIDRWEPELNEAIPNDERDTTPVAMATTLRKLTLGELLTTL 176
DB 137 QLOQTCTQLOATAEPSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLTRDLKCV 195
OY 177 ASROOLI 183
DB 196 ANGDOCV 202

RESULT 3

US-10-142-717-12
; Sequence 12, Application US/10142717
; Publication No. US20030104579A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Baum, Peter R.
; APPLICANT: Mosley, Bruce A.
; APPLICANT: Ketchum, Randal R.
; APPLICANT: Taylor, Scott L.
; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-142-717-12

Query Match
Best Local Similarity 5.6%; Score 75; DB 15; Length 202;
Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LILGAVLSRIDAGQGLRRIRHYSOND--LVEYSPTTEKHL-----TDGMTVREL----- 97
DB 17 LLLAAVLTQTQADPVPRAIRLPEVKAQCHIAQFKSLSPRELOAFKKAQDAIEKRLERDL 76
OY 98 -CSA-----AITMS--DNTAANLLTTTIGGPKELTAFLAN 129
DB 77 RCSSHLFPRAWDKQLOQVERPKALQAEVALTLKWMWTDALATITIGQPLHTLSHTS 136
OY 130 MGDHYT-----RIDRWEPELNEAIPNDERDTTPVAMATTLRKLTLGELLTTL 176
DB 137 QLOQTCTQLOATAEPSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLTRDLKCV 195
OY 177 ASROOLI 183
DB 196 ANGDOCV 202

RESULT 4

US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Atmout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-8

Query Match
Best Local Similarity 5.2%; Score 70.5; DB 11; Length 193;
Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

OY 59 IDAGQGLRRIRHYSONDLYESP--VTEKHLTDGMTVRELCSAATMSDNTAANLLTTT 116
DB 36 IGPQIOVG-----VOYGEDVYHERHLNDYRSVKOVVEAASHIEOR----- 77
OY 117 IGGPKELTAFLHNNGDHYTRIDRWE-----PELNEAIPNDERD 154
DB 78 -GGTETRTAF-----GIFARSEAFQKGRGAKKVMVITIDGSHSDPLEKAYIOOSERD 132
OY 155 TTPPVAMA 162
DB 133 NTKRYAVA 140

RESULT 5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds
(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_157T

Perfect score: 1348
Sequence: 1 HPEFLVKVADBDQIGARVG.....TMDERNRQIAEIGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA.*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	457	33.9	94	US-10-102-806-760	Sequence 760, App
2	75	5.6	202	US-10-127-816-9	Sequence 9, Appl
3	75	5.6	202	US-10-142-717-12	Sequence 12, Appl
4	70.5	5.2	193	US-09-805-354-8	Sequence 8, Appl
5	70.5	5.2	193	US-10-144-259-8	Sequence 8, Appl
6	70	5.2	206	US-10-156-761-13867	Sequence 13867, A
7	69.5	5.2	184	US-10-101-464-765	Sequence 765, App
8	69	5.1	149	US-10-156-761-8136	Sequence 8136, Ap
9	67.5	5.0	149	US-10-233-926-4	Sequence 4, Appl
10	67.5	5.0	187	US-09-948-018-8	Sequence 8, Appl
11	67	5.0	178	US-09-738-626-5680	Sequence 5680, Ap
12	67	5.0	194	US-10-156-761-8623	Sequence 8623, Ap
13	66	4.9	202	US-10-127-816-11	Sequence 11, Appl
14	64.5	4.8	191	US-10-156-761-8978	Sequence 8978, Ap
15	64.5	4.8	195	US-10-156-761-12656	Sequence 12656, A

16	64	4.7	162	US-09-738-626-4796	Sequence 4796, Ap
17	64	4.7	179	US-09-764-868-757	Sequence 757, App
18	64	4.7	179	US-10-106-698-4858	Sequence 4858, Ap
19	64	4.7	202	US-10-189-346-12	Sequence 12, Appl
20	63.5	4.7	88	US-10-100-252-6	Sequence 4, Appl
21	63.5	4.7	144	US-10-131-406-4	Sequence 4, Appl
22	63.5	4.7	189	US-09-815-242-5862	Sequence 5862, Ap
23	63.5	4.7	189	US-09-815-242-12979	Sequence 12979, A
24	63.5	4.7	206	US-09-738-626-5425	Sequence 5425, Ap
25	62.5	4.6	146	US-10-101-464-515	Sequence 515, App
26	62.5	4.6	190	US-10-156-761-9507	Sequence 9507, Ap
27	62.5	4.6	191	US-09-764-870-390	Sequence 390, App
28	62.5	4.6	191	US-10-125-540-350	Sequence 350, App
29	62	4.6	174	US-09-864-761-35777	Sequence 35777, A
30	62	4.6	191	US-09-927-827-64	Sequence 64, Appl
31	62	4.6	202	US-10-156-761-13460	Sequence 13460, A
32	61.5	4.6	142	US-09-862-027-31	Sequence 31, Appl
33	61.5	4.6	196	US-09-738-626-6624	Sequence 6624, Ap
34	61.5	4.6	200	US-10-156-761-14086	Sequence 12086, A
35	61	4.5	189	US-09-764-870-523	Sequence 523, App
36	61	4.5	189	US-10-125-540-523	Sequence 523, App
37	60.5	4.5	160	US-09-882-227-230	Sequence 230, App
38	60.5	4.5	162	US-10-141-531-5	Sequence 5, Appl
39	60.5	4.5	202	US-10-189-346-16	Sequence 16, Appl
40	60	4.5	190	US-09-738-626-4637	Sequence 4637, Ap
41	59.5	4.4	182	US-10-156-761-13056	Sequence 12056, A
42	59.5	4.4	196	US-10-219-220-265	Sequence 265, App
43	59.5	4.4	206	US-09-741-669-461	Sequence 461, App
44	59.5	4.4	206	US-09-815-242-10335	Sequence 10335, A
45	59	4.4	111	US-09-867-550-1868	Sequence 1868, Ap

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins and Antibodies
; FILE REFERENCE: P103P1c1
; CURRENT APPLICATION NUMBER: US/10/102,806
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-760

Query Match 33.9% Score 457; DB 15; Length 94;
Best Local Similarity 97.8%; Pred. No. 8.5e-40;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 75 NDIVEYSPVTEKHITDGMTVRELCSAITSMDNTANMLITTTIGSPRELFALHNMGDHV 134
DB 1 NDIVEYSPVTEKHITDGMTVRELCSAITSMDNTANMLITTTIGSPRELFALHNMGDHV 60

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22681
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22681

Query Match 4.6%; Score 62; DB 4; Length 132;
Best Local Similarity 29.8%; Pred. No. 46;
Matches 28; Conservative 8; Mismatches 34; Indels 24; Gaps 4;

OY 144 LINEATPNDERTTVAATITRLITGELLTLASRQOLIMWEADKYAGPLLRSAIP-- 201
DB 34 LROGIPN--HAGRRQAATVLPMEPGE-----RRVAFPMLKSPPPPR 74
OY 202 AGWFIADKSGAGERSGIIA--ALGPDGKPSR 232
DB 75 AAPEYADRAAGPPPTGTCACAPRRADAGVYPAR 108

RESULT 14
US-09-345-473E-31
; Sequence 31, Application US/09345473E
; Patent No. 6558903
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin
; TITLE OF INVENTION: NO. 6558903el Kinases and Uses Thereof
; FILE REFERENCE: 35800/183781
; CURRENT APPLICATION NUMBER: US/09/345,473E
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-345-473E-31

Query Match 4.6%; Score 61.5; DB 4; Length 142;
Best Local Similarity 41.7%; Pred. No. 58;
Matches 15; Conservative 5; Mismatches 11; Indels 5; Gaps 1;
OY 210 SGAGERSGRIIALGP-----DGKPSRIIVYITG 240
DB 47 SGSGSRDSGGTIALAPELLNDGKASKADSVISFG 82

RESULT 15
US-09-252-991A-24737
; Sequence 24737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24737
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24737

Query Match 4.6%; Score 61.5; DB 4; Length 160;

Best Local Similarity 37.5%; Pred. No. 70;
Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

OY 3 ETLVKVDAEDQICARVGYIELDINSGRII-----ESFRPERPPMNS 45
DB 109 EILKRVKSFILKSOGAAIVHLEVDENNPRAYSFYKSGFEFRERFVMS 156

Search completed: September 10, 2003, 12:31:03
Job time : 11.8571 secs

	Query Match	4.8%:	Score 64.5:	DB 4:	Length 178;
	Best Local Similarity	26.8%:	Pred. NO. 38;		
	Matches 37;	Conservative 14;	Mismatches 49;	Indels 39;	Gaps 8;
OY	110 ANLLTTIG---GPKELTA-----LHNNGDVTRLDRWEPFLNEAIPDEBDRTTPVAMA	162			
	:				
Dd	25 ARVLVELAQRRRGELAAHNGSLAHNHGCDIVAR-----AGEDLP--RDVDGLGAAA	74			
OY	163 TTLKLLITGELL-----TLASRQQLIDMMEADKVAGPLLRSLPAGWFADRSKG	211			
Dd	75 LRHHLYTAGALOGENHEEGIGDTGAAGQAQAVVG--QDQEVAGPRV--GLQAGLFPLMYCGD	130			
OY	212 AGERGRSGIITAALGPDKR	229			
	:				
Dd	131 A-----LVGVGGQAGR	141			

RESULT 9
US-08-624-677A-2
; Sequence 2, Application US/08624677A
; Patent No. 6476192
; GENERAL INFORMATION:
; APPLICANT: Lally, Nicola C.
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Dubey, Jitender P.
; TITLE OF INVENTION: Antigens useful for the Serodiagnosis of
; TITLE OF INVENTION: Neosporosis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Graeter, Janelle S.
; STREET: Rm. 411, Bldg. 005, BANC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/624,677A
; FILING DATE: 15-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.

[illegible]

```

RESULT 10
US-08-225-480-4
; Sequence 4, Application US/08225480
; Patent No. 5817793
GENERAL INFORMATION:
APPLICANT: LEVY, STUART B.
TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
TITLE OF INVENTION: ASSAYS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,085
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: T0359/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-225-480-4

```


Db 12 MGDGIGERGRGTGNTGCPRLRVYVPGDESPITLGRASPEDLIT--NLSPITMHPPS 69
QY 181 OLIDWMEA-DKVGAPL-----RSALPAGMFLADSGAGERSRGITIALGDPGRSR 232
Db 70 RDDDWVEPFDALGTAFTDGTGKATPAVGGI-----GASTRGS-GILASLSPROPAR 123

RESULT 2
US-09-252-991A-24241

Sequence 24241, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24241
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match

Best Local Similarity 23.6%; Score 72.5; DB 4; Length 197;
Matches 37; Conservative 23; Mismatches 42; Indels 55; Gaps 7;

QY 94 VIELCSAATMSDNTAANILLTTIGPKELAFLEHMGDHYRLDRWEPELNEAIPNDER 153
Db 12 YAKLSAARTSAASLAASVPLIPATP--TSARFSAGASL-----PSFYI 55
QY 154 DTTTPVA-MATLRLKLTGELLTLA-----SQOQLDMWEADKVAQP----- 194
Db 56 ATTSPLACRACRRLSCG--LARANSTPGSTSHRR--WSSNSISAPVSAGRSMPI 110
QY 195 -----LRSALPAGMFLADSGAGERSG 216
Db 111 SMAPARAIVTWPVITFATPAANWHSATNATASERG 147

RESULT 3

US-08-858-207A-519

Sequence 519, Application US/08858207A
Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Robert
TITLE OF INVENTION: NO. 6348328el Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Glumil, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 519:

SEQUENCE CHARACTERISTICS:
LENGTH: 170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-519

Query Match

Best Local Similarity 21.6%; Score 71.5; DB 4; Length 170;
Matches 46; Conservative 34; Mismatches 70; Indels 63; Gaps 8;

QY 15 LGARVYIELDLNSGKILSFREPERPMSTFVLGAVLSRIDAGQ--EQLGRRIHY 72
Db 1 LRKNIGLVLDLSIRELSPISCTKKSVMRFR-----AAAVVDASFIQELPQYD 54
QY 73 SQNDLVSPYTEHLTDGKMTVRELCSAATMSDNTAANILLTTIGPKELTAFLEHMGD 132
Db 55 -----SPVSEKSSSTGQQLAPARTVASO-----PKILI----- 86
QY 133 HYRLDRWEPELNEAIPNDERDTTPYAMATTLRLKLTGELLTLASQOQLDMWEADKVA 192
Db 87 -----LDEATNIDETESLV--QASLAKMNGR-TTIAHRLSTIODANCI- 131
QY 193 GPLRSALPAGMFLADSGAGERSRGITIALG 225
Db 132 -----YVLDKGRITIEGTHBELALG 152

RESULT 4

US-08-991-890-4

Sequence 4, Application US/08991890
Patent No. 6114307

GENERAL INFORMATION:

APPLICANT: Jaspers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_157T
Perfect score: 1348
Sequence: 1 HPETLVKYNDAEDLGARVG.....TMDERNQIAEIGASLIRHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74.5	5.5	124	US-08-311-731A-202	Sequence 202, App
2	72.5	5.4	197	US-09-252-991A-24241	Sequence 24241, A
3	71.5	5.3	170	US-08-858-207A-519	Sequence 519, App
4	71	5.3	159	US-08-991-890-4	Sequence 4, Appl1
5	69	5.1	169	US-09-328-352-4491	Sequence 4491, Ap
6	65.5	4.9	158	US-09-010-809-19	Sequence 19, Appl
7	65	4.8	150	US-09-239-909-2	Sequence 2, Appl1
8	64.5	4.8	178	US-09-252-991A-29942	Sequence 29942, A
9	64	4.7	203	US-08-624-677A-2	Sequence 2, Appl1
10	63.5	4.7	144	US-08-225-480-4	Sequence 4, Appl1
11	63.5	4.7	144	US-09-118-445-4	Sequence 4, Appl1
12	62.5	4.6	198	US-09-413-814-87	Sequence 87, Appl
13	62	4.6	132	US-09-252-991A-22681	Sequence 22681, A
14	61.5	4.6	142	US-09-345-473E-31	Sequence 31, Appl
15	61.5	4.6	160	US-09-252-991A-24737	Sequence 24737, A
16	61	4.5	181	US-08-482-142-195	Sequence 195, App
17	61	4.5	181	US-08-478-572-195	Sequence 195, App
18	61	4.5	181	US-08-484-296-195	Sequence 195, App
19	61	4.5	189	US-08-671-548C-48	Sequence 48, Appl
20	60.5	4.5	103	US-09-732-210-1282	Sequence 1282, Ap
21	60.5	4.5	167	US-08-690-849-2	Sequence 2, Appl1
22	60.5	4.5	167	US-09-004-053-2	Sequence 2, Appl1
23	60	4.5	144	US-09-252-991A-25578	Sequence 25578, A
24	59.5	4.4	144	US-09-134-001C-5269	Sequence 5269, Ap
25	59.5	4.4	178	US-09-220-711-24	Sequence 24, Appl
26	59.5	4.4	178	US-09-242-999-24	Sequence 24, Appl
27	59	4.4	49	US-08-284-465-9	Sequence 9, Appl1

28	59	4.4	145	US-09-134-001C-4982	Sequence 4982, Ap
29	59	4.4	171	US-09-107-532A-3979	Sequence 3979, Ap
30	59	4.4	174	US-08-557-122A-6	Sequence 6, Appl1
31	59	4.4	174	US-09-262-666-5	Sequence 5, Appl1
32	59	4.4	180	US-08-194-905-5	Sequence 5, Appl1
33	59	4.4	199	US-09-252-991A-30363	Sequence 30363, A
34	59	4.4	200	US-08-557-122A-12	Sequence 12, Appl
35	59	4.4	200	US-09-262-666-12	Sequence 12, Appl
36	58.5	4.3	167	US-09-328-352-6278	Sequence 6278, Ap
37	58.5	4.3	103	US-09-252-991A-24921	Sequence 24921, A
38	58	4.3	136	US-09-252-991A-30622	Sequence 30622, A
39	58	4.3	168	US-08-451-947-6	Sequence 6, Appl1
40	58	4.3	168	US-08-424-826A-6	Sequence 6, Appl1
41	58	4.3	168	US-08-928-694-6	Sequence 6, Appl1
42	58	4.3	168	US-08-450-842-6	Sequence 6, Appl1
43	58	4.3	168	US-08-451-390-6	Sequence 6, Appl1
44	58	4.3	168	PCT-US91-06950-6	Sequence 6, Appl1
45	58	4.3	183	US-09-252-991A-31347	Sequence 31347, A

ALIGNMENTS

RESULT 1
US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LARVAE FOR
DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
MOLECULE TYPE: linear
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202
Query Match 5.5%; Score 74.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 1.7;
Matches 34; Conservative 19; Mismatches 23; Gaps 8;
QY 130 MGDHV--TRLDWRPELNN---EAIIPNDRDT---TFVAMATRLRLTGLGELLTLASRO 180

THIS PAGE BLANK (USPTO)

XX Sequence 158 AA:

Query Match 6.0%; Score 80.5; DB 22; Length 158;
Best Local Similarity 33.3%; Pred. No. 4.1;
Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

QY 110 ANLLITTTGGPKELTAFLHNMGDVTRLDREPELNEAIPNDESDTTTPVAMATTLR--K 167
| : : : | | | : | | | : | | | : | | | : | | | : | | | : | | | :
DB 80 ARMATSTSGAPKVSASHSLTARGASPWTYGPWDEPSTRPAI-----TTSAPLAFPDTLRSGR 134

QY 168 LVTGELLTLASRQOLIDW 185
| : : : | | | : | | | : | | | : | | | : | | | : | | | :
DB 135 GRMGSWYLLATRIPIRLW 152

Search completed: September 10, 2003, 12:24:51
Job time : 32.7143 secs

Dn 6Z DFKGSEKXETQ-----AMLS-----NIVLPT---PSOLDREVLANML 99

Oy 133 DHVTRLDHWEPELNPAINDERDTPPVAAATPLRKLT-----GELTTLASROOLIDWME 187
: : : : :
Db 100 ERIEOLO-----ELNE-----NRRAIIVIRBMTIPTILNEROALIEFIK 138
: : : : :
Oy 188 ----ADKVASPLLRGAL 200
: : : : :
Db 139 ENNPESDRIT--LLESSL 153

RESULT 15
AAU45138
ID AAU45138 standard; Protein: 158 AA.
XX AAU45138;
XX
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #6034.
DE SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
RW dermatological; osteopathic; neuroprotectant.
XX
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
PD 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Alonsoeuvre J, Zhang Y, Jen S, Carter D;
XX DR N-PSDB; AASS9525.
XX WPI; 2001-616774/71.

Propionibacterium acnes polypeptides and nucleic acids useful for vaccinating against and diagnosing infections, especially useful for treating acne vulgaris -

Example 1; SEQ ID No 6333; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at fnr.wipo.int/pub/wipubled

